

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 29, 2004, 05:02:07 ; Search time 7419 Seconds

(without alignments)
10936.536 Million cell updates/sec

Title: US-09-787-491B-18

Perfect score: 1872

Sequence: 1 gccctctgctgctgtgtgtc.....tatcaaaaaaaaaaaaaa 1872

Scoring table:

OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 3470272 seqs, 21671516995 residues

Word size : 0

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

```
1: GenEmbl:*
2: gb_ba:*
3: gb_hkg:*
4: gb_in:*
5: gb_om:*
6: gb_ov:*
7: gb_pat:*
8: gb_ph:*
9: gb_pl:*
10: gb_pt:*
11: gb_ro:*
12: gb_sts:*
13: gb_sy:*
14: gb_un:*
15: gb_vi:*
16: em_ba:*
17: em_fun:*
18: em_hum:*
19: em_in:*
20: em_mu:*
21: em_om:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_hkg_hum:*
31: em_hkg_inv:*
32: em_hkg_other:*
33: em_hkg_mus:*
34: em_hkg_pih:*
35: em_hkg_rtd:*
36: em_hkg_mam:*
37: em_hkg_vrt:*
38: em_sy:*
39: em_hkg_hum:*
40: em_hkg_mus:*
41: em_hkg_other:*
```

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1872	100.0	1802	6	BD223765 Human RNA
2	1720	91.9	3502	6	AF419331 Homo sapi
3	1483	79.2	1836	6	AX876414 Sequence
4	1483	79.2	1836	6	BD156116 Primer fo
5	1483	79.2	1836	6	AK001286 Homo sapi
6	1439	76.9	1814	2	AY007101 Homo sapi
7	1439	76.9	199002	2	AC090439 Homo sapi
8	1276	68.2	16745	9	AY048592 Homo sapi
9	1276	68.2	106433	9	EX511012 Homo sapi
10	1276	68.2	173307	9	AL590609 Homo sapi
11	1273	68.0	2073	9	HSMB801468 Human DNA
12	1225	65.4	2834	9	AK093226 Homo sapi
13	940	50.2	1326	6	BD223773 Homo sapi
14	786	42.0	2473	6	AX713349 Human RNA
15	786	42.0	2473	6	AK054635 Homo sapi
16	726	38.8	874	9	BC001107 Homo sapi
17	674	36.0	2247	6	AX876834 Homo sapi
18	674	36.0	2247	6	BD156341 Homo sapi
19	674	36.0	2247	6	AK001479 Homo sapi
20	618	33.0	739	9	AY150181 Homo sapi
21	531	28.4	1633	9	BC005039 Homo sapi
22	520	27.8	1058	9	BC010074 Homo sapi
23	513	27.4	2940	9	AF419332 Homo sapi
24	509	27.2	702	6	AX868774 Sequence
25	509	27.2	702	6	BD148836 Primer fo
26	509	27.2	2918	6	AX877218 Sequence
27	509	27.2	2918	6	BD156546 Primer fo
28	509	27.2	2918	9	AK001656 Homo sapi
29	496	26.5	620	9	AF047448 Homo sapi
30	496	26.5	620	9	AF067730 Homo sapi
31	481	25.7	1556	9	AY150180 Homo sapi
32	467	24.9	789	9	AF448427 Homo sapi
33	430	23.0	737	6	BD020635 Novel gen
34	430	23.0	737	6	BD100573 Novel gen
35	420	22.4	881	6	AX867568 Sequence
36	420	22.4	881	6	BD147630 Primer fo
37	394	21.0	882	6	AX868161 Sequence
38	394	21.0	882	6	BD148223 Primer fo
39	370	19.8	392	6	AX886243 Sequence
40	370	19.8	392	6	BD025853 Sequence
41	297	15.9	2562	9	AK055311 Homo sapi
42	297	15.9	3957	9	AK125834 Homo sapi
43	265	14.2	485	6	AR414810 Sequence
44	265	14.2	485	6	BD110363 EST and e
45	247	13.2	247	6	AX913051 Sequence

ALIGNMENTS

RESULT 1
LOCUS BD223765 1872 bp DNA linear PAT 17-JUL-2003
DEFINITION Human RNA-associated proteins.
ACCESSION BD223765
VERSION BD223765.1 GI:33033535
KEYWORDS JP 2002523045-A/1.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
1 (bases 1 to 1872)
Hillman,J.L., Yue,H., Tang,Y.T., Corley,N.C., Giegler,K.J.,
Gorgone,G.A., Patterson,C., Baughn,M.R., Ial,P., Bandman,O.,
Reddy,R., Azimzai,Y., Shih,L.L., Yang,J. and Lu,D.A.M.

TITLE Human RNA-associated proteins
JOURNAL Patent: JP 2002523045-A 1 30-Jul-2002;
INCYTE PHARMACEUTICALS INC
OS Homo sapiens (human)
PN JP 2002523045-A/1
PD 30-Jul-2002
PF 20-AUG-1999 JP 2000566425
PR 21-AUG-1998 US 60/097550, 12-JAN-1999 US 60/115639 PI
JENNIFER L HILLMAN, HENRY YUE, Y TOM TANG, NEIL C CORLEY, KARL J PI
GUEGLER,
PI GINA A GORGONE, CHANDRA PATTERSON, MARIAN R BAUGHN, PREETI LAL,
PI OLGA BANDMAN,
PI ROOPA REDDY, YALIDA AZIMZAI, LEO L SHIH, JUNMING
YANG, DYUNG AINA M
PI LU
PC C12N15/09, A61K38/00, A61K38/55, A61K45/00, A61P1/00, A61P1/04, PC
A61P1/16,
PC A61P1/18, A61P3/00, A61P3/10, A61P7/00, A61P7/06, A61P9/00, A61P11/
PC 00, A61P11/06,
PC A61P13/12, A61P17/00, A61P17/06, A61P19/00, A61P19/02, A61P19/06,
PC A61P19/10,
PC A61P21/00, A61P21/04, A61P29/00, A61P29/00, A61P31/00, A61P31/04,
PC A61P31/10,
PC A61P31/12, A61P31/18, A61P31/22, A61P35/00, A61P35/02, A61P37/00,
PC A61P37/08,
PC C07K14/47, C07K16/18, C12N1/15, C12N1/19, C12N1/21, C12N15/00,
PC C12P21/02, C12Q1/68,
PC G01N33/15, G01N33/50, G01N33/53, G01N33/566//C12P21/08, C12N15/00,
PC C12N5/00,
PC A61K37/02, A61K37/64
CC Incyte Identification No.: 399781CBI
FH Key
FT source 1. .1872 Location/Qualifiers
FT 1. .1872 Location/Qualifiers
FT /organism='Homo sapiens (human)'.
FT /organism="Homo sapiens"
FT /mol_type="genomic DNA"
FT /db_xref="taxon:9606"
ORIGIN
Query Match 100.0%; Score 1872; DB 6; Length 1872;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1872; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GCCCTCTACGCTGTGTCTGAGGCTCGGCCCTGAGCCCGGAGCGGTTTCTGAGCC 60
DB 1 GCCCTCTACGCTGTGTCTGAGGCTCGGCCCTGAGCCCGGAGCGGTTTCTGAGCC 60
QY 61 CGTTAGTGGCCCGGCGGAGACACGCGCGCCGCTATCCGCTACCTGGGCTCCCGCAAC 120
DB 61 CGTTAGTGGCCCGGCGGAGACACGCGCGCCGCTATCCGCTACCTGGGCTCCCGCAAC 120
QY 121 AGCTCTCTGCTGTCGAGAACGTCGCCGACGACACAGAGTCTGAACACTTGGGGCTGAA 180
DB 121 AGCTCTCTGCTGTCGAGAACGTCGCCGACGACACAGAGTCTGAACACTTGGGGCTGAA 180
QY 181 TTTGGTGGTATGTCCTATAGTTGATGTGATGTTCTCACTTGATTTCTACATCGCCGT 240
DB 181 TTTGGTGGTATGTCCTATAGTTGATGTGATGTTCTCACTTGATTTCTACATCGCCGT 240
QY 241 CCAAGAGATTTGCTATGTTCAATTTGAGAGTTCGTGATGTCGTAAGCGCTTTACAT 300
DB 241 CCAAGAGATTTGCTATGTTCAATTTGAGAGTTCGTGATGTCGTAAGCGCTTTACAT 300
QY 301 AATTGGACGAAAGTGTGTCGACGCGCAGATTGAATAACAGTTTGGCCGAGGGGAT 360
DB 301 AATTGGACGAAAGTGTGTCGACGCGCAGATTGAATAACAGTTTGGCCGAGGGGAT 360
QY 361 CGAAGACACCAATCATGATGAAGCCAGAGGAGGAGATGTGTCAGTTCTTACAGC 420
DB 361 CGAAGACACCAATCATGATGAAGCCAGAGGAGGAGATGTGTCAGTTCTTACAGC 420

QY 421 TATGATGATTATGACAGATACAGACTGTTCTAGAACCGGAAGTTATGAAGGAGAGATCA 480
DB 421 TATGATGATTATGACAGATACAGACTGTTCTAGAACCGGAAGTTATGAAGGAGAGATCA 480
QY 481 AGAAGTGGTCTTTTGATTTACACTATAGAAAGATCGTATAGTCTTAGAANAAGTAGACCG 540
DB 481 AGAAGTGGTCTTTTGATTTACACTATAGAAAGATCGTATAGTCTTAGAANAAGTAGACCG 540
QY 541 ACTGGAAGACACGCGGTAGAGAGAGCCATTCCGACAAATGATAGACCAAACTGACAGTGA 600
DB 541 ACTGGAAGACACGCGGTAGAGAGAGCCATTCCGACAAATGATAGACCAAACTGACAGTGA 600
QY 601 ATACCCAGTACAGTCTGCTTACTACACTTCAAGAAAGATCTGAAAGCGGAAAAAGAAC 660
DB 601 ATACCCAGTACAGTCTGCTTACTACACTTCAAGAAAGATCTGAAAGCGGAAAAAGAAC 660
QY 661 AAAGAGGGCAGTTCAAGCCGCAAAAGGTGGGTGGAAGGGTCCGCAATAGATACGCT 720
DB 661 AAAGAGGGCAGTTCAAGCCGCAAAAGGTGGGTGGAAGGGTCCGCAATAGATACGCT 720
QY 721 ACGAATATTTTGACCTGCTGCTGGAAGAAAGATTAAGATGTTATCGAAAACTACATGAAT 780
DB 721 ACGAATATTTTGACCTGCTGCTGGAAGAAAGATTAAGATGTTATCGAAAACTACATGAAT 780
QY 781 AATTGAAGTCCCTTCAAGTTTGAAGATGATTTAGAGCAATTAAGAAATTCAC 840
DB 781 AATTGAAGTCCCTTCAAGTTTGAAGATGATTTAGAGCAATTAAGAAATTCAC 840
QY 841 TTTGTAAGTGGGAAACCTAATCCCTAATATGAATAGGTTTATATGTTCAATGGGTAC 900
DB 841 TTTGTAAGTGGGAAACCTAATCCCTAATATGAATAGGTTTATATGTTCAATGGGTAC 900
QY 901 AGGTCCATTAATTAATTTATGGAACCTAGATGCTGAATATCAAGAGACAGCCATAGT 960
DB 901 AGGTCCATTAATTAATTTATGGAACCTAGATGCTGAATATCAAGAGACAGCCATAGT 960
QY 961 CTCTTACAGTGCCTGTGTGTCTGTCTGCAAACTGAATTTGGGTGGAAAAAGTATGTCC 1020
DB 961 CTCTTACAGTGCCTGTGTGTGTCTGTCTGCAAACTGAATTTGGGTGGAAAAAGTATGTCC 1020
QY 1021 AATATTAAGATTCATTTTGGCATTTATGGAATCTGCAATATCAAGAGACAGCCATAGT 1080
DB 1021 AATATTAAGATTCATTTTGGCATTTATGGAATCTGCAATATCAAGAGACAGCCATAGT 1080
QY 1081 AGTGTTCCTGCTTATATCAATTTGTTGTGCACTGCTGTTATTTACTTGAACACAC 1140
DB 1081 AGTGTTCCTGCTTATATCAATTTGTTGTGCACTGCTGTTATTTACTTGAACACAC 1140
QY 1141 ATGCAATTAATCTGCTTAACTACTCTCTCCAGATTAATTCATTAATTTGACAT 1200
DB 1141 ATGCAATTAATCTGCTTAACTACTCTCTCCAGATTAATTCATTAATTTGACAT 1200
QY 1201 CCAAGTAAAGAGGCCCATCTCTTCCAGATTAATTCATTAATTTGACAT 1260
DB 1201 CCAAGTAAAGAGGCCCATCTCTTCCAGATTAATTCATTAATTTGACAT 1260
QY 1261 TATGAGCCCTTAAGTGGGCAATCATTTGATGATTAATGAAGAAATAGATAATTC 1320
DB 1261 TATGAGCCCTTAAGTGGGCAATCATTTGATGATTAATGAAGAAATAGATAATTC 1320
QY 1321 CCTTATTCAGTAATGTCCTACGAGACATCTAGGAATCATTAAGTAAGTGGCCCAT 1380
DB 1321 CCTTATTCAGTAATGTCCTACGAGACATCTAGGAATCATTAAGTAAGTGGCCCAT 1380
QY 1381 GTTTTGTGAGGTGTTATTCATTAATTAATTTTACACATTCGATCAATGATTA 1440
DB 1381 GTTTTGTGAGGTGTTATTCATTAATTAATTTTACACATTCGATCAATGATTA 1440
QY 1441 TAGAACAATAATACATCAAGATTAAGTAATTTGTTATCTGCAATTTAAAGTAT 1500
DB 1441 TAGAACAATAATACATCAAGATTAAGTAATTTGTTATCTGCAATTTAAAGTAT 1500
QY 1501 CAGATATTTGATCACTATTAATTAATGAAGAAATGATTTAATCTGTAATTAACCTGG 1560

Db 1501 CCAGTATTGATCACTTATTTAAATATATGAAAAATGATTAATCTGTAAATAAAGCTGG 1560
Qy 1561 TTTATTTGTCAGTGCATGATTAATACTAGATTAATAATAATTTGTTACTGCGCCAC 1620
Db 1561 TTTATTTGTCAGTGCATGATTAATACTAGATTAATAATAATTTGTTACTGCGCCAC 1620
Qy 1621 AAACACATCTCTGATATATACCCCAAAATTAAGTATTAAGTTGATAGGTATTAAG 1680
Db 1621 AAACACATCTCTGATATATACCCCAAAATTAAGTATTAAGTTGATAGGTATTAAG 1680
Qy 1681 AGACTGGGTCTATATTAATTTAGGACAGACAGAGCTGTATCTTAATCTGAT 1740
Db 1681 AGACTGGGTCTATATTAATTTAGGACAGACAGAGCTGTATCTTAATCTGAT 1740
Qy 1741 TTAGTATGTTCTGTAATGAGAAAAAGTTCACTTAATTAATTTTGTAGTTATCAATG 1800
Db 1741 TTAGTATGTTCTGTAATGAGAAAAAGTTCACTTAATTAATTTTGTAGTTATCAATG 1800
Qy 1801 TACATTTTATAGGGGACATGTTCTGTATAGCGAATAATTAATTAATTAATCAAAA 1860
Db 1801 TACATTTTATAGGGGACATGTTCTGTATAGCGAATAATTAATTAATTAATCAAAA 1860
Qy 1861 AAAAAAAAAA 1872
Db 1861 AAAAAAAAAA 1872

RESULT 2
AF419331 3502 bp mRNA linear PRI 17-OCT-2001
LOCUS AF419331
DEFINITION Homo sapiens TLS-associated protein TASR-1 mRNA, complete cds,
alternative transcript.
ACCESSION AF419331
VERSION AF419331.1 GI:16265856
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens (human)
REFERENCE Bukacinska, Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
AUTHORS Clinton, J.M., Chansky, H.A., Odell, D.D., Zielinska-Kwiatkowska, A.
and Yang, L.
TITLE Characterization and expression of the human gene encoding two
JOURNAL TLS-associated serine-arginine (TASR) proteins
REFERENCE Unpublished
AUTHORS 2 (bases 1 to 3502)
TITLE Direct Submission
JOURNAL Submitted (14-SEP-2001) Department of Orthopedics, University of
Washington, 1660 S. Columbian Way, GMR 151, Seattle, WA 98108, USA
FEATURES
source
1..3502
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/chromosome="1"
/map="NT_004359"
/cell_line="Hela"
77..628
/note="splicing factor"
/codon_start=1
/product="TLS-associated protein TASR-1"
/protein_id="A116665.1"
/db_xref="GI:16265857"
/translation="MSRYLRPPTSLFVNVAADDTSEDIIRFEFGYPIVDVYELD
FYTRPRGPAIVQFEDVRDAEDALHNDKRWICGQIEIQAGDKKTPVOMAKKGR
NVYSSRDYDIDYRRSRSSRYSRSPDNYNRRSSPNSPTGPRSRSSHS
DNDRPNCWMTQYSSAYYSRKI"
1819..1824
/note="alternative"
polYA_signal
ORIGIN

Query Match 91.9%; Score 1720; DB 9; Length 3502;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1840; Conservative 0; Mismatches 0; Indels 1; Gaps 1;
Qy 18 GTCTGAGGCTGCGCCGCTGAGCCGCGAGACGGTTTCTAGAGCCGTTAGTGCGCCGCG 77
Db 1 GTCTGAGGCTGCGCCGCTGAGCCGCGAGACGGTTTCTAGAGCCGTTAGTGCGCCGCG 60
Qy 78 GAGACAGCCCGCCGATGTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 137
Db 61 GAGACAGCCCGCCGATGTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 120
Qy 138 GAACTGGCCGACGACACAGGCTGGAAGCTGGGGGTGAATTTGGTCTTATGATGTC 197
Db 121 GAACTGGCCGACGACACAGGCTGGAAGCTGGGGGTGAATTTGGTCTTATGATGTC 180
Qy 198 TATAGTGTATGATGTTCCACTGATTTCTACACTGCGCTCCAGAGAGATTTGCTTA 257
Db 181 TATAGTGTATGATGTTCCACTGATTTCTACACTGCGCTCCAGAGAGATTTGCTTA 240
Qy 258 TGTTCATTTGAGATGTTGCTGATGCTGAAGACGCTTTACATATTTGACAGAAAGTG 317
Db 241 TGTTCATTTGAGATGTTGCTGATGCTGAAGACGCTTTACATATTTGACAGAAAGTG 300
Qy 318 GATTGTGACGCGAGATTTGAATACGTTTGGCCAGGGGGATGGAAGACACCAATCA 377
Db 301 GATTGTGACGCGAGATTTGAATACGTTTGGCCAGGGGGATGGAAGACACCAATCA 360
Qy 378 GATGAAAGCCAGGAGGAGAGAAATGTGTCAGTTCTTCAAGCTATGATGATTAAGACAG 437
Db 361 GATGAAAGCCAGGAGGAGAGAAATGTGTCAGTTCTTCAAGCTATGATGATTAAGACAG 420
Qy 438 ATACAGACGTTCTGAAAGCCGAGTTATGAAGAGAGATCAAGAACTCGCTTTTGA 497
Db 421 ATACAGACGTTCTGAAAGCCGAGTTATGAAGAGAGATCAAGAACTCGCTTTTGA 480
Qy 498 TTACACATATGAGATGTTATGCTTCTAGAAACATGACCCGATCGAAGACACAGCGG 557
Db 481 TTACACATATGAGATGTTATGCTTCTAGAAACATGACCCGATCGAAGACACAGCGG 540
Qy 558 TAG-AGAAGCATTTCCGACATGATAGACCAATCTGAGCTGGAATCCCGATGACGTC 616
Db 541 TAGAGAGAGCATTTCCGACATGATAGACCAATCTGAGCTGGAATCCCGATGACGTC 600
Qy 617 TGCTTACTACATTTCAAGAAAGATCTGAAGCGGAAAAAGAACCAAGAAAGGCGATTC 676
Db 601 TGCTTACTACATTTCAAGAAAGATCTGAAGCGGAAAAAGAACCAAGAAAGGCGATTC 660
Qy 677 AGCAGCAAGGGGTGGAGGTGCTGAGATGTAATCTGACGAATTTTGTACTC 736
Db 661 AGCAGCAAGGGGTGGAGGTGCTGAGATGTAATCTGACGAATTTTGTACTC 720
Qy 737 TGCTTGAAGAAAGATTAAGAAAGATGTTATCGAAAACTACATGGAATTAAGAGTCCCTCA 796
Db 721 TGCTTGAAGAAAGATTAAGAAAGATGTTATCGAAAACTACATGGAATTAAGAGTCCCTCA 780
Qy 797 AGTTTGAAGATAGCATTTTGAACAATAAAGAAATTCATTTGTACTGTGAGAA 856
Db 781 AGTTTGAAGATAGCATTTTGAACAATAAAGAAATTCATTTGTACTGTGAGAA 840
Qy 857 CTATCCCTTAATATATGATGTTTATATGATTCATGGGTAAACAGTCCATATAATAT 916
Db 841 CTATCCCTTAATATATGATGTTTATATGATTCATGGGTAAACAGTCCATATAATAT 900
Qy 917 ATTGAAACTAGAGATGCTGAATATATCAAGGACAGCATAGTCTTCAAGTCCCTCT 976
Db 901 ATTGAAACTAGAGATGCTGAATATATCAAGGACAGCATATGCTTCAAGTCCCTCT 960
Qy 977 GTTGTGCTGCTCAAACTGAATTTGGGTGAGAAAGGATATGTCATATTAAGATTCAT 1036
Db 961 GTTGTGCTGCTCAAACTGAATTTGGGTGAGAAAGGATATGTCATATTAAGATTCAT 1020

QY 1037 TTTTGCATTAATTGGCAAAATCTTGCTTTGTTTATTTTGGTCCAGTGTTCCTGTTAA 1096
 DB 1021 TTTTGCATTAATTGGCAAAATCTTGCTTTGTTTATTTTGGTCCAGTGTTCCTGTTAA 1080
 QY 1097 TCAATGCTTGGTGGCAATCTGTTTATTTACTGTGACACACAGTGTTCATCTG 1156
 DB 1081 TCAATGCTTGGTGGCAATCTGTTTATTTACTGTGACACACAGTGTTCATCTG 1140
 QY 1157 TCTTAATCTCTCTCTCCAGGTAAATTCAAATTATTTGACATCCAGCTAAGAGGCCC 1216
 DB 1141 TCTTAATCTCTCTCTCCAGGTAAATTCAAATTATTTGACATCCAGCTAAGAGGCCC 1200
 QY 1217 ATCTCTTCACTCTTTTCCCTAGTATGATATTCGCAAAATTTTATTTAGAGCCCTTACTG 1276
 DB 1201 ATCTCTTCACTCTTTTCCCTAGTATGATATTCGCAAAATTTTATTTAGAGCCCTTACTG 1260
 QY 1277 TGGGCAATCAATCTGATGATATTTGAGAAAAATAGATAATTCCTTATTCAGTAATG 1336
 DB 1261 TGGGCAATCAATCTGATGATATTTGAGAAAAATAGATAATTCCTTATTCAGTAATG 1320
 QY 1337 TCTACTGACACAACTATAGTAAATCATTAACAGTATGCTCATTTGTTTGAAGTGT 1396
 DB 1321 TCTACTGACACAACTATAGTAAATCATTAACAGTATGCTCATTTGTTTGAAGTGT 1380
 QY 1397 GTTATTCATTAACAATTTTATTAACCATTCGTATTCATATTAATTAAGAACCAATATCG 1456
 DB 1381 GTTATTCATTAACAATTTTATTAACCATTCGTATTCATATTAATTAAGAACCAATATCG 1440
 QY 1457 ATCAAGATTAAGTATTTGTTGTTATCTGCATTTAAAGATCCAGATTTGATCA 1516
 DB 1441 ATCAAGATTAAGTATTTGTTGTTATCTGCATTTAAAGATCCAGATTTGATCA 1500
 QY 1517 TTATTAATAATAAGAAAAATGATTTAATCTGTAATAAACTGTTTATTTGACGTGAC 1576
 DB 1501 TTATTAATAATAAGAAAAATGATTTAATCTGTAATAAACTGTTTATTTGACGTGAC 1560
 QY 1577 TGTATTAATCTAGTATTAATAATTTGTTTCTGCTGCTCACAAACACATGAGAGAT 1636
 DB 1561 TGTATTAATCTAGTATTAATAATTTGTTTCTGCTGCTCACAAACACATGAGAGAT 1620
 QY 1637 ATTAACCCCAAAATAGTATTTAATCTGATTAAGATTAAGAGACTGGGTGCTATA 1696
 DB 1621 ATTAACCCCAAAATAGTATTTAATCTGATTAAGATTAAGAGACTGGGTGCTATA 1680
 QY 1697 TTAGATTAATTTGAGGACAGACAGAGCTGTATTCCTAACTGATTTATGATGTTCTGTA 1756
 DB 1681 TTAGATTAATTTGAGGACAGACAGAGCTGTATTCCTAACTGATTTATGATGTTCTGTA 1740
 QY 1757 TTGAAGAAATGTTTACCAATTTACTTTTGTAGTATTAACATGATCATTTATAGGGGA 1816
 DB 1741 TTGAAGAAATGTTTACCAATTTACTTTTGTAGTATTAACATGATCATTTATAGGGGA 1800
 QY 1817 CATGCTCTGTATGACGATAAATACTTTTATAGTATCA 1857
 DB 1801 CATGCTCTGTATGACGATAAATACTTTTATAGTATCA 1841
 RESULT 3
 AX876414 1836 bp DNA linear PAT 17-DRC-2003
 LOCUS AX876414
 DEFINITION Sequence 11319 from Patent EP1074617.
 ACCESSION AX876414
 VERSION AX876414.1 GI:40031150
 KEYWORDS
 SOURCE
 ORGANISM
 Homo sapiens (human)
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE
 AUTHORS Ota, T., Isogai, T., Nishikawa, T., Hayashi, K., Saito, K., Yamamoto, J.,
 Ishii, S., Sugiyama, T., Wakamatsu, A., Nagai, K. and Otsuki, T.
 TITLE Primers for synthesizing full-length cDNA and their use
 JOURNAL Patent: EP 1074617-A 11319 07-FEB-2001;

FEATURES
 source
 CDS
 ORIGIN
 Query Match 79.2%; Score 1483; DB 6; Length 1836;
 Best Local Similarity 99.8%; Pred. No. 0;
 Matches 1833; Conservative 0; Mismatches 1; Indels 3; Gaps 2;
 Research Association for Biotechnology (JRP)
 Location/Qualifiers
 1..1836
 /organism="Homo sapiens"
 /mol_type="unassigned DNA"
 /db_xref="taxon:9606"
 73..624
 /note="unnamed protein product"
 /codon_start=1
 /protein_id="CAE89263.1"
 /db_xref="GI:40031151"
 /translation="MSRYLRPNTSLFVRNVAADTRESDELRFRRGRPIVDVYPLD
 FYTRPRGFAYVCEFDVDAEDALHNDKWI CGRQLEIQAGDRKIPNMKAEGR
 NVSSRYDDYDRYRSRSEYERRSRFSFDYNYRYSYSPRNSRPLGRFRSRSHS
 DNDPNCNMNYDSAYTSRKI"
 22 GAGGCTCGGCGGCTGAGCCGCGACCGGTTGCTGAGCCGCTTGTGCGCCGCGGCGA 81
 DB 1 GAGGCTCGGCGGCTGAGCCGCGACCGGTTGCTGAGCCGCTTGTGCGCGCGA 60
 QY 82 CACGCGCGCGCATGTCCGCTACCTGCGTCCCCCAACAGGCTCTGTGTCAGAAAC 141
 DB 61 CACGCGCGCGCATGTCCGCTACCTGCGTCCCCCAACAGGCTCTGTGTCAGAAAC 120
 QY 142 GTGCGCGACACACACAGGCTGTGAAGACTTGGCGGCTGAAATTTGTGCTTATAT 201
 DB 121 GTGCGCGACACACACAGGCTGTGAAGACTTGGCGGCTGAAATTTGTGCTTATAT 180
 QY 202 GTTGAATGTGATGTTCACCTTGAATTTTACACTGCGCGCTCAAGAGATTTGCTATAT 261
 DB 181 GTTGAATGTGATGTTCACCTTGAATTTTACACTGCGCGCTCAAGAGATTTGCTATAT 240
 QY 262 CAATTTGAGAGATGTCGTGATGCTGAAGACGCTTTACATTAATTTGACAGAAAGTGA 321
 DB 241 CAATTTGAGAGATGTCGTGATGCTGAAGACGCTTTACATTAATTTGACAGAAAGTGA 300
 QY 322 TGTGACGCGCAGATTGAATACAGTTTGTCCGACGGGGGATCGAAAGACCAATACAGAT 381
 DB 301 TGTGACGCGCAGATTGAATACAGTTTGTCCGACGGGGGATCGAAAGACCAATACAGAT 360
 QY 382 AAAGCCAGAGAGAGAGAAATGTATACAGTTTGTCAAGTATGATGATATGACAGATAC 441
 DB 361 AAAGCCAGAGAGAGAGAAATGTATACAGTTTGTCAAGTATGATGATATGACAGATAC 420
 QY 442 ACACGTTCTAGAAGCCGAAGTATGAAGAGAGAGATCAAGAAAGTGTCTTTGATTAAC 501
 DB 421 ACACGTTCTAGAAGCCGAAGTATGAAGAGAGAGATCAAGAAAGTGTCTTTGATTAAC 480
 QY 502 AACATATGAAGATCGTATATGCTCTGAGAAACAGTATGACCACTGAGATTAATGACAGATAC 560
 DB 481 AACATATGAAGATCGTATATGCTCTGAGAAACAGTATGACCACTGAGATTAATGACAGATAC 540
 QY 561 AGAAGCATTCGACACATATGACCAATCTGAGCTGAGATATCCAGTACAGTTCTGCT 620
 DB 541 AGAAGCATTCGACACATATGACCAATCTGAGCTGAGATATCCAGTACAGTTCTGCT 600
 QY 621 TACTACACTTCAAGAAAGTCTGAAGCCGGAACCAAGCAAGAGAGGAGTTCAGAGC 680
 DB 601 TACTACACTTCAAGAAAGTCTGAAGCCGGAACCAAGCAAGAGAGGAGTTCAGAGC 660
 QY 681 ACCAAAGGCTGGTGAAGGCTGCTGCAATGATATCTGACGATTAATTTGACTCTGCT 740
 DB 661 ACCAAAGGCTGGTGAAGGCTGCTGCAATGATATCTGACGATTAATTTGACTCTGCT 720
 QY 741 CTGAAGAAATGAAGATGTTATCGAAACTACATGGAATTAATGAAGTCCCTCAAGT 800
 DB 721 CTGAAGAAATGAAGATGTTATCGAAACTACATGGAATTAATGAAGTCCCTCAAGT 780

```

QY 801 TGAAGTAAGCATTTTAGGCAAAATAAAAGAAATTCACCTTTGTACTTGTGGAAACTTA 860
DB 781 TTTAAGTAAGCATTTTAGGCAAAATAAAAGAAATTCACCTTTGTACTTGTGGAAACTTA 840
QY 861 TCCCTAAATATGAAATAGTTTATTTGATTCAATGGGTAAACAGTCCATTAATAATATG 920
DB 841 TCCCTAAATATGAAATAGTTTATTTGATTCAATGGGTAAACAGTCCATTAATAATATG 900
QY 921 GAAACTAGAGTGTCTGAATATTCAGAGAAACAGCCATAGTCTCTTACAGTGGCTGTG 980
DB 901 GAAACTAGAGTGTCTGAATATTCAGAGAAACAGCCATAGTCTCTTACAGTGGCTGTG 960
QY 981 GTCGTCTCAACCTGAATTTGGTGGGAAAAGGTATGTCATATPAAAAGTCCATTTT 1040
DB 961 GTCTGTCTCAACCTGAATTTGGTGGGAAAAGGTATGTCATATPAAAAGTCCATTTT 1020
QY 1041 GCCATTAATTTGGCAATCTTGCCTTTGTTATTTTGGTCCAGTGTCTCTGTTAAAT 1100
DB 1021 GCCATTAATTTGGCAATCTTGCCTTTGTTATTTTGGTCCAGTGTCTCTGTTAAAT 1080
QY 1101 TTGCTTTGTGGCATCTGTGTTTATTTTACCTGTACACCAATGCAGTTTACATCTGCTT 1160
DB 1081 TTGCTTTGTGGCATCTGTG--TATTCTGTACACCAATGCAGTTTACATCTGCTT 1138
QY 1161 AACTACTCTCCGAGGTAATTCCTCAATTAATTTGATTCAGCTAAAGAGGCCCATCT 1220
DB 1139 AACTACTCTCTCCGAGGTAATTCCTCAATTAATTTGATTCAGCTAAAGAGGCCCATCT 1198
QY 1221 CTCTCACTCTTCTTCTAGTCAGTAATTCAGCAAAATTTATTTAGAGCCCTTACGTGGG 1280
DB 1199 CTCTCACTCTTCTTCTAGTCAGTAATTCAGCAAAATTTATTTAGAGCCCTTACGTGGG 1258
QY 1281 CAAATCATTTGTAAGTATTTGAGAAAATAGATAATTCCTTATTCAGTAATATGCTA 1340
DB 1259 CAAATCATTTGTAAGTATTTGAGAAAATAGATAATTCCTTATTCAGTAATATGCTA 1318
QY 1341 CTGAGCACAATCTAGTAATTCAGTAATTCAGTATGAGCCCTATTTGTTGAGGTGTGTA 1400
DB 1319 CTGAGCACAATCTAGTAATTCAGTAATTCAGTATGAGCCCTATTTGTTGAGGTGTGTA 1378
QY 1401 TTCAATCAATATTTTACACCATTCGTATCAATTAATTTAGAACCAATATACGATCA 1460
DB 1379 TTCAATCAATATTTTACACCATTCGTATCAATTAATTTAGAACCAATATACGATCA 1438
QY 1461 AGGATTAAGTAATTTGTTATCTGCAATTTAAAGATTCAGTATTTGATCAATTAAT 1520
DB 1439 AGGATTAAGTAATTTGTTATCTGCAATTTAAAGATTCAGTATTTGATCAATTAAT 1498
QY 1521 TATAATTAATGAAATTAATGATTTAATCTGTAATTAATCTGTTATTTGCAAGTCTGA 1580
DB 1499 TATAATTAATGAAATTAATGATTTAATCTGTAATTAATCTGTTATTTGCAAGTCTGA 1558
QY 1581 ATATACTAGATTAATTAATGATTTAATCTGCTCCTCACCAACAATGCTAGGATATA 1640
DB 1559 ATATACTAGATTAATTAATGATTTAATCTGCTCCTCACCAACAATGCTAGGATATA 1618
QY 1641 CCCCCCAATTAATGATTTAATCTTGCATTAAGTAATTAAGGAGACTGGTGTCTAATTA 1700
DB 1619 CCCCCCAATTAATGATTTAATCTTGCATTAAGTAATTAAGGAGACTGGTGTCTAATTA 1678
QY 1701 ATATTTTGAAGAGAGAGAGAGCTGTATTCCTAATGATTTAGTATGTTCTGTAATTA 1760
DB 1679 ATATTTTGAAGAGAGAGAGAGCTGTATTCCTAATGATTTAGTATGTTCTGTAATTA 1738
QY 1761 GAAATATGTTACCAATTAATCTTTTATGATTTACATGATCAATTTTATAGGGAGCATG 1820
DB 1739 GAAATATGTTACCAATTAATCTTTTATGATTTACATGATCAATTTTATAGGGAGCATG 1798
QY 1821 TTTCTGTATAGGAATTAATTAATCTTTTATAGTATCA 1857
DB 1799 TTTCTGTATAGGAATTAATTAATCTTTTATAGTATCA 1835

```

```

RESULT 4
BD156116
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
PD 09-JUL-2002
PF 28-JUL-2000
PI TOSHIO OTA,TAKAO ISOGAI,TEISUO NISHIKAWA,KOJI HAYASHI,KAORU
PI SAITO,
PI JUNICHI YAMAMOTO,SHIZUKO ISHII,TOMOYASU SUGIYAMA,AI WAKAMATSU,
PI KEIICHI NAGAI,TEISUJI OTSUKI
PC
C12N15/09,C07K14/47,C07K16/18,C12N1/15,C12N1/19,C12N1/21,C12N5/
10',C12P21/02,C1201/68//C12P21/08,G06F17/30,C12N15/00,C12N5/00
Primer for synthesizing full-length cDNA and use thereof FH Key
Location/Qualifiers
(73)..(621).
FT CDS
1..1836
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
Query Match 79.2%; Score 1483; DB 6; Length 1836;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 1833; Conservative 0; Mismatches 1; Indels 3; Gaps 2;
ORIGIN
22 GAGGCTCGGCGGCTGAGCGGCGGAGCGTTTGTGAGCCCGGTTAGTGGCCCGCGAG 81
DB 1 GAGGCTCGGCGGCTGAGCGGCGGAGCGTTTGTGAGCCCGGTTAGTGGCCCGCGAG 60
QY 82 CACGCCGCGCATGTCCGCTACCTGCGTCCCCCAACACGCTCTGTTGTCAGGAGC 141
DB 61 CACGCCGCGCATGTCCGCTACCTGCGTCCCCCAACACGCTCTGTTGTCAGGAGC 120
QY 142 GTGGCCGAGCACACCGGTCTGAAGACTTGGCGCGGTGAATTTGTCGTTATGCTATA 201
DB 121 GTGGCCGAGCACACCGGTCTGAAGACTTGGCGCGGTGAATTTGTCGTTATGCTATA 180
QY 202 GTTGATGTGATTTTCACTGATTTCTACCTGCGCGCTGCAAGAGATTTGCTATGTT 261
DB 181 GTTGATGTGATTTTCACTGATTTCTACCTGCGCGCTGCAAGAGATTTGCTATGTT 240
QY 262 CAATTTGAGAGTGTGATGCTGAAGACGCTTTACATATTTTGACAGAAAGTGAT 321
DB 241 CAATTTGAGAGTGTGATGCTGAAGACGCTTTACATATTTTGACAGAAAGTGAT 300
QY 322 TGTGACGGGCAATTAATTAAGTTTGGCCAGGGGAGTCGAAGACACCAATCGATG 381
DB 301 TGTGACGGGCAATTAATTAAGTTTGGCCAGGGGAGTCGAAGACACCAATCGATG 360
QY 382 AAAGCCAAAGAGAGAGAGATGTCAGTTCCTACAGCTATGATATTAAGACAGATAC 441
DB 361 AAAGCCAAAGAGAGAGAGATGTCAGTTCCTACAGCTATGATATTAAGACAGATAC 420
QY 442 AGACGTTTGAAGCCGAAATTATGAAGAGAGATCAAGAGTGGCTTTTGATTAC 501

```

Db AGACGTTCTAGAAAGCCGAACTTATGAAAGAGAGATCAAGAGTCGGTCTTTTGATTAAC 480
 QY AACATAGAGATGCTATTAAGTCTAGAAAACAGTAGACCGACTGGAGACCAAGCGCTAG- 560
 502 AACATAGAGATGCTATTAAGTCTAGAAAACAGTAGACCGACTGGAGACCAAGCGCTAG- 560
 Db AACATAGAGATGCTATTAAGTCTAGAAAACAGTAGACCGACTGGAGACCAAGCGCTAG- 540
 481 AACATAGAGATGCTATTAAGTCTAGAAAACAGTAGACCGACTGGAGACCAAGCGCTAG- 540
 QY AGAAGCAATCCGACCAATGATAGACCAAACTGACGCTGGAAATACCCAGTACAGTTCGTCT 620
 561 AGAAGCAATCCGACCAATGATAGACCAAACTGACGCTGGAAATACCCAGTACAGTTCGTCT 620
 Db AGAAGCAATCCGACCAATGATAGACCAAACTGACGCTGGAAATACCCAGTACAGTTCGTCT 600
 541 AGAAGCAATCCGACCAATGATAGACCAAACTGACGCTGGAAATACCCAGTACAGTTCGTCT 600
 QY TACATACCTTCAAAAAGATCTGAAAGCGGAAAAAGAACCAAGAGGGCAGTTCAAGCG 680
 621 TACATACCTTCAAAAAGATCTGAAAGCGGAAAAAGAACCAAGAGGGCAGTTCAAGCG 680
 Db TACATACCTTCAAAAAGATCTGAAAGCGGAAAAAGAACCAAGAGGGCAGTTCAAGCG 660
 601 TACATACCTTCAAAAAGATCTGAAAGCGGAAAAAGAACCAAGAGGGCAGTTCAAGCG 660
 QY ACCAAAGGGTGGGAGAGGCTGCTGCAATGAAATCTGTAAGATATTTGACTCTGGT 740
 681 ACCAAAGGGTGGGAGAGGCTGCTGCAATGAAATCTGTAAGATATTTGACTCTGGT 740
 Db ACCAAAGGGTGGGAGAGGCTGCTGCAATGAAATCTGTAAGATATTTGACTCTGGT 720
 661 ACCAAAGGGTGGGAGAGGCTGCTGCAATGAAATCTGTAAGATATTTGACTCTGGT 720
 QY CTGAAAAATAAAGATGTTATGCAAAATCAAGAAATTAAGTCCCTTCAAGTT 800
 741 CTGAAAAATAAAGATGTTATGCAAAATCAAGAAATTAAGTCCCTTCAAGTT 800
 Db CTGAAAAATAAAGATGTTATGCAAAATCAAGAAATTAAGTCCCTTCAAGTT 780
 721 CTGAAAAATAAAGATGTTATGCAAAATCAAGAAATTAAGTCCCTTCAAGTT 780
 QY TGAAGATAGAGATTTAGCAAAATTAAGGAAATTCACCTTGTACTTGGAAACTAA 860
 801 TGAAGATAGAGATTTAGCAAAATTAAGGAAATTCACCTTGTACTTGGAAACTAA 860
 Db TGAAGATAGAGATTTAGCAAAATTAAGGAAATTCACCTTGTACTTGGAAACTAA 840
 781 TGAAGATAGAGATTTAGCAAAATTAAGGAAATTCACCTTGTACTTGGAAACTAA 840
 QY TCCCTAAATATGATAGGTTTATATGATTCATGAGGTAACAGGTCCATTAATTAATTG 920
 861 TCCCTAAATATGATAGGTTTATATGATTCATGAGGTAACAGGTCCATTAATTAATTG 920
 Db TCCCTAAATATGATAGGTTTATATGATTCATGAGGTAACAGGTCCATTAATTAATTG 900
 841 TCCCTAAATATGATAGGTTTATATGATTCATGAGGTAACAGGTCCATTAATTAATTG 900
 QY GAAATCTAGAGATCTGGAATATCAAGAAACAGCCATGCTCTTCAAGTGCCTGTTG 980
 921 GAAATCTAGAGATCTGGAATATCAAGAAACAGCCATGCTCTTCAAGTGCCTGTTG 980
 Db GAAATCTAGAGATCTGGAATATCAAGAAACAGCCATGCTCTTCAAGTGCCTGTTG 960
 901 GAAATCTAGAGATCTGGAATATCAAGAAACAGCCATGCTCTTCAAGTGCCTGTTG 960
 QY GTCTGCTCAAACTGAAATGAGTGGGAAAAAGGATGCTCAATATTAAGTTCATTTT 1040
 981 GTCTGCTCAAACTGAAATGAGTGGGAAAAAGGATGCTCAATATTAAGTTCATTTT 1040
 Db GTCTGCTCAAACTGAAATGAGTGGGAAAAAGGATGCTCAATATTAAGTTCATTTT 1020
 961 GTCTGCTCAAACTGAAATGAGTGGGAAAAAGGATGCTCAATATTAAGTTCATTTT 1020
 QY GGCATTAATGGAAGATCTTGGCTTTGTTATTTGGTGCAGAGTTCCTGCTTAATCA 1080
 1041 GGCATTAATGGAAGATCTTGGCTTTGTTATTTGGTGCAGAGTTCCTGCTTAATCA 1080
 Db GGCATTAATGGAAGATCTTGGCTTTGTTATTTGGTGCAGAGTTCCTGCTTAATCA 1060
 1021 GGCATTAATGGAAGATCTTGGCTTTGTTATTTGGTGCAGAGTTCCTGCTTAATCA 1060
 QY TTGCTTTGTTGGCAGTCTGCTTTATTTATTTGATACCAAGCAAGAGTTCATCTGCTT 1160
 1101 TTGCTTTGTTGGCAGTCTGCTTTATTTATTTGATACCAAGCAAGAGTTCATCTGCTT 1160
 Db TTGCTTTGTTGGCAGTCTGCTTTATTTATTTGATACCAAGCAAGAGTTCATCTGCTT 1140
 1081 TTGCTTTGTTGGCAGTCTGCTTTATTTATTTGATACCAAGCAAGAGTTCATCTGCTT 1140
 QY AACCTACTCTCCGCCAGGTAATTCGAATATTTGACATCCAGCTAAGAGGCCCATCT 1220
 1161 AACCTACTCTCCGCCAGGTAATTCGAATATTTGACATCCAGCTAAGAGGCCCATCT 1220
 Db AACCTACTCTCCGCCAGGTAATTCGAATATTTGACATCCAGCTAAGAGGCCCATCT 1200
 1139 AACCTACTCTCCGCCAGGTAATTCGAATATTTGACATCCAGCTAAGAGGCCCATCT 1200
 QY CTTTCTACCTCTTTCTAGTCAATATTTGACAAATTTATTTGAGCCCTTACTGTGGG 1280
 1221 CTTTCTACCTCTTTCTAGTCAATATTTGACAAATTTATTTGAGCCCTTACTGTGGG 1280
 Db CTTTCTACCTCTTTCTAGTCAATATTTGACAAATTTATTTGAGCCCTTACTGTGGG 1260
 1199 CTTTCTACCTCTTTCTAGTCAATATTTGACAAATTTATTTGAGCCCTTACTGTGGG 1260
 QY CAATATATGTAATGTAATTAAGAAAAATGATATTTCCCTTATTCAGTAATGTCTA 1340
 1281 CAATATATGTAATGTAATTAAGAAAAATGATATTTCCCTTATTCAGTAATGTCTA 1340
 Db CAATATATGTAATGTAATTAAGAAAAATGATATTTCCCTTATTCAGTAATGTCTA 1320
 1259 CAATATATGTAATGTAATTAAGAAAAATGATATTTCCCTTATTCAGTAATGTCTA 1320
 QY CTGAGACCAATCTAGGAAATCATTAAGATAGGCTCATTTGTTTGGAGGTGTGTA 1400
 1341 CTGAGACCAATCTAGGAAATCATTAAGATAGGCTCATTTGTTTGGAGGTGTGTA 1400
 Db CTGAGACCAATCTAGGAAATCATTAAGATAGGCTCATTTGTTTGGAGGTGTGTA 1380
 1319 CTGAGACCAATCTAGGAAATCATTAAGATAGGCTCATTTGTTTGGAGGTGTGTA 1380
 QY TTGATTAACAATATTTTAACCAATTCGATCAATGTAATTAAGAACCAATTAACATCA 1460
 1401 TTGATTAACAATATTTTAACCAATTCGATCAATGTAATTAAGAACCAATTAACATCA 1460
 Db TTGATTAACAATATTTTAACCAATTCGATCAATGTAATTAAGAACCAATTAACATCA 1440
 1379 TTGATTAACAATATTTTAACCAATTCGATCAATGTAATTAAGAACCAATTAACATCA 1440
 QY AGAATAGTAATGTAATGTAATTCGCAATTTAAAGATCCAGTATTTGATCAATTA 1520
 1461 AGAATAGTAATGTAATGTAATTCGCAATTTAAAGATCCAGTATTTGATCAATTA 1520
 Db AGAATAGTAATGTAATGTAATTCGCAATTTAAAGATCCAGTATTTGATCAATTA 1500
 1439 AGAATAGTAATGTAATGTAATTCGCAATTTAAAGATCCAGTATTTGATCAATTA 1500
 QY TATTAATTAATGAAAAATGATTTATCTGTAATTAACCTGTTTATGTCAGTGA 1580
 1521 TATTAATTAATGAAAAATGATTTATCTGTAATTAACCTGTTTATGTCAGTGA 1580
 Db TATTAATTAATGAAAAATGATTTATCTGTAATTAACCTGTTTATGTCAGTGA 1560
 1499 TATTAATTAATGAAAAATGATTTATCTGTAATTAACCTGTTTATGTCAGTGA 1560

QY ATATATTAGATTAATTAATTAATTTTACTCTGCTGACCAACACACATGCTAGATATA 1640
 1581 ATATATTAGATTAATTAATTAATTTTACTCTGCTGACCAACACACATGCTAGATATA 1640
 Db ATATATTAGATTAATTAATTAATTTTACTCTGCTGACCAACACACATGCTAGATATA 1620
 1559 ATATATTAGATTAATTAATTAATTTTACTCTGCTGACCAACACACATGCTAGATATA 1620
 QY CCCCCAAATTAAGATTTAATTTGATTTAGTATTAAGAGAGAGTGGTGTCTAATTAATG 1700
 1641 CCCCCAAATTAAGATTTAATTTGATTTAGTATTAAGAGAGAGTGGTGTCTAATTAATG 1700
 Db CCCCCAAATTAAGATTTAATTTGATTTAGTATTAAGAGAGAGTGGTGTCTAATTAATG 1680
 1619 CCCCCAAATTAAGATTTAATTTGATTTAGTATTAAGAGAGAGTGGTGTCTAATTAATG 1680
 QY ATTATTTGAGGAGACAGAGAGCTGTTATCTTAATGATTTAGTATTTAGTATTTGA 1760
 1701 ATTATTTGAGGAGACAGAGAGCTGTTATCTTAATGATTTAGTATTTAGTATTTGA 1760
 Db ATTATTTGAGGAGACAGAGAGCTGTTATCTTAATGATTTAGTATTTAGTATTTGA 1740
 1679 ATTATTTGAGGAGACAGAGAGCTGTTATCTTAATGATTTAGTATTTAGTATTTGA 1740
 QY GAAATGTTACCAAAATTTATCTTTTATGATTTAGTATTTAGTATTTAGTATTTGA 1820
 1761 GAAATGTTACCAAAATTTATCTTTTATGATTTAGTATTTAGTATTTAGTATTTGA 1820
 Db GAAATGTTACCAAAATTTATCTTTTATGATTTAGTATTTAGTATTTAGTATTTGA 1800
 1739 GAAATGTTACCAAAATTTATCTTTTATGATTTAGTATTTAGTATTTAGTATTTGA 1800
 QY TTCTGTATAGCGAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATCA 1857
 1821 TTCTGTATAGCGAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATCA 1857
 Db TTCTGTATAGCGAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATCA 1835
 1799 TTCTGTATAGCGAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATCA 1835

RESULT 5
 AK001286
 LOCUS
 DEFINITION
 Homo sapiens TLS-associated protein TASR-2 mRNA.
 ACCESSION
 AK001286
 VERSION
 AK001286.1 GI:7022445
 KEYWORDS
 oligo cloning; fls (full insert sequence).
 SOURCE
 Homo sapiens (human)
 ORGANISM
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE
 AUTHORS
 1
 Isogai, T., Oca, T., Hayashi, K., Sugiyama, T., Otsuki, T., Suzuki, Y.,
 Nishikawa, T., Nagai, K., Sugano, S., Shiratori, A., Sudo, H.,
 Wadatsuna, M., Hosokawa, T., Kaku, Y., Kodaira, H., Kondo, H.,
 Sugawara, M., Takahashi, M., Chiba, Y., Ishida, S., Murakawa, K.,
 Ono, Y., Takiguchi, S., Matanabe, S., Kimura, K., Murakami, K.,
 Ishii, S., Kawai, Y., Saito, K., Yamamoto, J., Wakamatsu, A.,
 Nakamura, Y., Nagahata, K., Masuno, Y., Ninomiya, K. and Iwayanagi, T.
 NEDO human cDNA sequencing project
 Unpublished
 2 (bases 1 to 1836)
 TITLE
 JOURNAL
 REFERENCE
 AUTHORS
 Isogai, T. and Otsuki, T.
 Direct Submission
 Submitted (16-FEB-2000) Takao Isogai, Helix Research Institute,
 Genomics Laboratory; 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
 (E-mail:genom@helix.co.jp, Tel:81-438-52-3975, Fax:81-438-52-3986)
 COMMENT
 NEDO human cDNA sequencing project supported by Ministry of
 International Trade and Industry for Biotechnology; cDNA library
 sequencing; Research Association for Biotechnology; cDNA library
 construction; 5'- & 3'-end one pass sequencing and clone selection;
 Helix Research Institute (supported by Japan Key Technology Center
 etc.) and Department of Virology, Institute of Medical Science,
 University of Tokyo.
 FEATURES
 Location/Qualifiers
 1..1836
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="NT2RP1000272"
 /cell_line="NT2"
 /cell_type="teratocarcinoma"
 /clone_11b="NT2RP1"
 /note="cloning vector: pUC19/PL3-mRNA from NT2 neuronal
 precursor cells after 48-hours retinoic acid (RA)
 induction."
 73..624
 /note="unnamed protein product"
 /codon_start=1
 /protein_id="BAA91601.1"

CDS

ORIGIN

/db xref="GI:7022446"
/translation="MSRYLRPNTSLFVNVDVTRSEDLRREFRGRYIVDVVPLD
FYTRPGRFAVQFEDVRDAEDALHNLKRMWICQRIEIQFGGRKTPNOKAKGR
NVYSSRRDDYDRYRSRYSRRSRSSRSPDYVRYSRSPRNSRPGRPBRSSHS
DMDRBNCSWNTGYSSAVYTSRKI"

Query Match 79.2%; Score 1483; DB 9; Length 1836;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 1833; Conservative 0; Mismatches 1; Indels 3; Gaps 2;

QY 22 GAGGCTCGGCGCGCTGAGCGCGGAGCGGTTGCTGAGCCCGTTAGTGCGCCCGCGGAGA 81
DB 1 GAGGCTCGGCGCGCTGAGCGCGGAGCGGTTGCTGAGCCCGTTAGTGCGCCCGCGGAGA 60
QY 82 CAGCGCGCGCGCATGTCGCGTACCTGCTGCCCAACACAGCTCTGTGTGCTGAGAAC 141
DB 61 CAGCGCGCGCGCATGTCGCGTACCTGCTGCCCAACACAGCTCTGTGTGCTGAGAAC 120
QY 142 GTGGCGGAGACACACAGGTCTGAAGAATTGGCGGCGTGAATTTGCTGATGCTCTATA 201
DB 121 GTGGCGGAGACACACAGGTCTGAAGAATTGGCGGCGTGAATTTGCTGATGCTCTATA 180
QY 202 GTTGAATGTATGTTCACTGATTTCTACACTGCGCGCTCCAGAGGATTTGCTTATGTT 261
DB 181 GTTGAATGTATGTTCACTGATTTCTACACTGCGCGCTCCAGAGGATTTGCTTATGTT 240
QY 262 CAATTGAGAGTGTGCTGATGCTGAGACGCTTACATTAATTTGAGACAGAAAGTGAAT 321
DB 241 CAATTGAGAGTGTGCTGATGCTGAGACGCTTACATTAATTTGAGACAGAAAGTGAAT 300
QY 322 TGTGACGCGGAGATTGAAATACAGTTTGCCCAAGGCGGATCGAAAGACCAAAATCAGATG 381
DB 301 TGTGACGCGGAGATTGAAATACAGTTTGCCCAAGGCGGATCGAAAGACCAAAATCAGATG 360
QY 382 AAAGCGAAGAGAGGAGATGTGACAGTCTTCAAGCTATGATGATTAAGACAGATAC 441
DB 361 AAAGCGAAGAGAGGAGATGTGACAGTCTTCAAGCTATGATGATTAAGACAGATAC 420
QY 442 AGACGTTCTAGAACCGGAGTTATGAAAGAGAGATCAAGAGTGGTCTTTTGAATAC 501
DB 421 AGACGTTCTAGAACCGGAGTTATGAAAGAGAGATCAAGAGTGGTCTTTTGAATAC 480
QY 502 AACTATAGAAATGTATATGCTTGAACAAGTACAGTGAAGCCAGCGGATG- 560
DB 481 AACTATAGAAATGTATATGCTTGAACAAGTACAGTGAAGCCAGCGGATG- 540
QY 561 AGAGCCATTCGACCAATGATGACCAAACTGCGAGCTGGAATACCCAGTACAGTTCGCT 620
DB 541 AGAGCCATTCGACCAATGATGACCAAACTGCGAGCTGGAATACCCAGTACAGTTCGCT 600
QY 621 TACTACATTCAGAAAGATCTGAAAGCGGAAAAAGAACCAAGAGGCGAGTTCAAGCG 680
DB 601 TACTACATTCAGAAAGATCTGAAAGCGGAAAAAGAACCAAGAGGCGAGTTCAAGCG 660
QY 681 ACCAAAGGCGGTGGAAGGCTGCTGAGTGAATGAATCTGTAAGAAATTTTGACTCGGT 740
DB 661 ACCAAAGGCGGTGGAAGGCTGCTGAGTGAATGAATCTGTAAGAAATTTTGACTCGGT 720
QY 741 CTGAAAGATGAAAGATGTTATGAAACCTACATGAAATGATGAGTCCCTCAAGTT 800
DB 721 CTGAAAGATGAAAGATGTTATGAAACCTACATGAAATGATGAGTCCCTCAAGTT 780
QY 801 TGAAGTAAAGATTTTGAACAATTAAGAAATTCACCTTGTACTGTGGAACCTAA 860
DB 781 TGAAGTAAAGATTTTGAACAATTAAGAAATTCACCTTGTACTGTGGAACCTAA 840
QY 861 TCCCTAAATATGATAGTGTATATGATTCATGCGGTAACAGGTCATTAATTAATATG 920
DB 841 TCCCTAAATATGATAGTGTATATGATTCATGCGGTAACAGGTCATTAATTAATATG 900
QY 921 GAAACTGAGATGTGTAATATCAAGGAAGACAGCAATGCTCTTACAGGCTCTGTG 980

DB 901 GAAACTGAGATGTGTAATATCAAGGAAGACAGCCATGATCTCTTACAGTGCCTGTG 960
QY 981 GTCTCTCAACCTGAATTTGGGTGGGAAAAGGTATGCTCAATATTAAGATTCATTTT 1040
DB 961 GTCTCTCAACCTGAATTTGGGTGGGAAAAGGTATGCTCAATATTAAGATTCATTTT 1020
QY 1041 GCCATTAATGGAATCTTGCCTTGTATATTTTGGGCGAGTGTCTGCTTATACAT 1100
DB 1021 GCCATTAATGGAATCTTGCCTTGTATATTTTGGGCGAGTGTCTGCTTATACAT 1080
QY 1101 TTGCTTGTGGCATCTGTGTATTTATCTGTACACACATGACAGTTTACATCTGTCT 1160
DB 1081 TTGCTTGTGGCATCTGTG--TATTTACTGTACACACATGACAGTTTACATCTGTCT 1138
QY 1161 AACTATCTCTTCCAGGTAATTTCAATTAATTTGACATCCAGCTAAGAGGCCCATCT 1220
DB 1139 AACTATCTCTTCCAGGTAATTTCAATTAATTTGACATCCAGCTAAGAGGCCCATCT 1198
QY 1221 CTTCACACTCTTTCCTGAGTATATTCAGCAAAATTTTATGAGGCCCTTACGTGGG 1280
DB 1199 CTTCACACTCTTTCCTGAGTATATTCAGCAAAATTTTATGAGGCCCTTACGTGGG 1258
QY 1281 CAATCATTTGATCTGATTAATGAAAAATAGATTAATTCCTTATTCAGTAAATGCTA 1340
DB 1259 CAATCATTTGATCTGATTAATGAAAAATAGATTAATTCCTTATTCAGTAAATGCTA 1318
QY 1341 CTGAGCAAACTCTGTGTAATGATTAACAGTATGCGCTCATTTGTTTGTGAGGTGTTA 1400
DB 1319 CTGAGCAAACTCTGTGTAATGATTAACAGTATGCGCTCATTTGTTTGTGAGGTGTTA 1378
QY 1401 TTCAATCAATATTTTACACCAATCGTATCAATGTAATTAAGACAATATACATCA 1460
DB 1379 TTCAATCAATATTTTACACCAATCGTATCAATGTAATTAAGACAATATACATCA 1438
QY 1461 AGATTAAGTAATTTGTGTGTTATGTCATTTAAAGATACAGTATTTGATCAATAT 1520
DB 1439 AGATTAAGTAATTTGTGTGTTATGTCATTTAAAGATTAAGATTAAGTATGATCAATAT 1498
QY 1521 TATTAATTAAGAAAAATGATTTAATCTGTATTAATCTGTTTATTTGCGAGTGTGA 1580
DB 1499 TATTAATTAAGAAAAATGATTTAATCTGTATTAATCTGTTTATTTGCGAGTGTGA 1558
QY 1581 AATATACAGTATTAATTAATTTGTTACTGTGCTCACCAACATGCTAGGATATA 1640
DB 1559 AATATACAGTATTAATTAATTTGTTACTGTGCTCACCAACATGCTAGGATATA 1618
QY 1641 CCCCCAAATTAAGTATTAATCTTTGCAATTAAGATTAAGAGACCTGGGTGCTAATATAG 1700
DB 1619 CCCCCAAATTAAGTATTAATCTTTGCAATTAAGATTAAGAGACCTGGGTGCTAATATAG 1678
QY 1701 ATTATTTTGGGAGACAGAGAGCTGTATCTTACCTGATGATTTAGTATGTTCTGTAATGGA 1760
DB 1679 ATTATTTTGGGAGACAGAGAGCTGTATCTTACCTGATGATTTAGTATGTTCTGTAATGGA 1738
QY 1761 GAAATGTTACCAAAATTAATTTAGTATTAATGATTAATGATTAATTTTGGGGAATG 1820
DB 1739 GAAATGTTACCAAAATTAATTTAGTATTAATGATTAATGATTAATTTTGGGGAATG 1798
QY 1821 TTCTGTGTATAGGAATTAATTAATTTTATATGATCA 1857
DB 1799 TTCTGTGTATAGGAATTAATTAATTTTATATGATCA 1835

RESULT 6
AY007101

LOCUS 1814 bp mRNA linear PRI 31-AUG-2000
DEFINITION Homo sapiens clone TCCCIA00269 mRNA sequence.
ACCESSION AY007101
VERSION AY007101.1 GI:9955992
KEYWORDS FLI CDNA.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Bukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
PUBMED
2 (bases 1 to 1814)
Yu, W., Andersson, B., Worley, K. C., Muzny, D. M., Ding, Y., Liu, W., Ricafrente, J. Y., Wentland, M. A., Lennon, G., and Gibbs, R. A. A 'double adaptor' method for improved shotgun library construction Anal. Biochem. 236 (1), 107-113 (1996)
6619474
8619474
JOURNAL
MEDLINE
PUBMED
3 (bases 1 to 1814)
Zhou, J., Yu, W., Tang, H., Mei, G., Tsang, Y. T. M., Bouck, J., Gibbs, R. A. and Margolin, J. F. Direct Submission Submitted (24-JUL-2000) Human Genome Sequencing Center and Texas Children's Cancer Center, Baylor College of Medicine, Houston, TX 77030, USA
COMMENT
The clone request should be directed to Dr. J. Margolin at Pediatrics-Hematology & Oncology, Texas Children's Center 102514, Houston, Texas 77030, USA. Telephone: 713-770-4583
margolin@bcm.tmc.edu.
FEATURES
source
1. .1814
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="HCC1A0269"
/clone_lib="constructed by Y.T.M. Tsang"
/dev_stage="infant"
/note="From patient with acute lymphoblastic leukemia, similar to Homo sapiens CDNA FLJ10424 fls, clone NT2RP1000272 with Genbank Accession Number AK001286"
ORIGIN
Query Match 76.9%; Score 1439; DB 9; Length 1814;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 1799; Conservative 0; Mismatches 0; Indels 4; Gaps 2;
53 GGTGAGCCGCTAGTGTGCGCCGCGGACGACGCGCCGCGCTACCTGCTGCGTGC 112
1 GGTGAGCCGCTAGTGTGCGCCGCGGACGACGCGCCGCGCTACCTGCTGCGTGC 60
113 CCCCCAACAGCTCTCTGTTGTCAGGAGCGTGCGCGACGACACAGGCTGGAAGCTTGC 172
61 CCCCCAACAGCTCTCTGTTGTCAGGAGCGTGCGCGACGACACAGGCTGGAAGCTTGC 120
173 GCGGTGAATTTGGTGGTATGTCCTATAGTTGATGTGATGTTCCACTTGAATTTCTACA 232
121 GCGGTGAATTTGGTGGTATGTCCTATAGTTGATGTGATGTTCCACTTGAATTTCTACA 180
233 CCGCGCGTCAAGAGATTGCTTATGTTCAATTTGAGATGTTGCTGATGCTGAAGACG 292
181 CTGCGCGTCAAGAGATTGCTTATGTTCAATTTGAGATGTTGCTGATGCTGAAGACG 240
293 CTTTACATATTTTGAACAGAAAGTGTGTTGAGCGCGAGATGTAATACAGTTTGGCC 352
241 CTTTACATATTTTGAACAGAAAGTGTGTTGAGCGCGAGATGTAATACAGTTTGGCC 300
353 AGGGGGATGGAAGACACCAATCAGATGAAGCCAGAGAGAGAGAGAGATGTCAGTT 412
301 AGGGGGATGGAAGACACCAATCAGATGAAGCCAGAGAGAGAGAGATGTCAGTT 360
413 CTTACAGCTATGATGATTTATGACAGATACAGAGCTTCTAGAACCCAGTTATGAAGA 472
361 CTTACAGCTATGATGATTTATGACAGATACAGAGCTTCTAGAACCCAGTTATGAAGA 420
473 GAGATCAAGAGATCGCTCTTTGATTACAACTATAGAGATGTTAGTCTGTAAGACA 532
421 GAGATCAAGAGATCGCTCTTTGATTACAACTATAGAGATGTTAGTCTGTAAGAAA-- 478

533 GTAGACCGACTGGAAGCCACGCGCTAG-AGAGCCATTCGACATGATGACCAACT 591
479 -TAGACGACTGGAAGCCACGCGCTAGACAGAGCCATTCGACATGATGACCAACT 537
592 GCGCTGGAATACCGAGTACAGTCTGCTTACTACATCTCAAGAAAGATCGAAAGCGGA 651
538 GCGCTGGAATACCGAGTACAGTCTGCTTACTACATCTCAAGAAAGATCGAAAGCGGA 597
652 AAAAGAACAAAGAAAGGCGAGTTCAGAGGACCAAGAGGTGGTGAAGGTGCTGAGAT 711
598 AAAAGAACAAAGAAAGGCGAGTTCAGAGGACCAAGAGGTGGTGAAGGTGCTGAGAT 657
712 GAATACGTAGGAATATTTTGAATCTGCTGAAAAGATAAGATGTATCGAAACT 771
658 GAATACGTAGGAATATTTTGAATCTGCTGAAAAGATAAGATGTATCGAAACT 717
772 ACATGGAATTAATGAAGTCCCTTCAAGTTTGAAGTAAGCATTTTGAACAAATTAAG 831
718 ACATGGAATTAATGAAGTCCCTTCAAGTTTGAAGTAAGCATTTTGAACAAATTAAG 777
832 AAATTCACCTTTGATCTTGTGGAACCTAATCCCTAATATGAATGTTTATGATTC 891
778 AAATTCACCTTTGATCTTGTGGAACCTAATCCCTAATATGAATGTTTATGATTC 837
892 ATGGGTACAGGTCATATAATAATTAATGAAGAACTAGAGATGTCGAATATCAAGAGAC 951
838 ATGGGTACAGGTCATATAATAATTAATGAAGAACTAGAGATGTCGAATATCAAGAGAC 897
952 AGCCATAGTCTTTCAGTGCCTGTTGTGCTGCTGCTCAAACTGAATGGTGGGAAAG 1011
898 AGCCATAGTCTTTCAGTGCCTGTTGTGCTGCTGCTCAAACTGAATGGTGGGAAAG 957
1012 GATAGTCCCAATTAAGTCCATTTTGCATATGAGCAAACTGCTGCTTAT 1071
958 GATAGTCCCAATTAAGTCCATTTTGCATATGAGCAAACTGCTGCTTAT 1017
1072 TTTGGTCCAGTGTTCGCTTATCAATTTGCTTGTGTCATCGTGTATTTACTT 1131
1018 TTTGGTCCAGTGTTCGCTTATCAATTTGCTTGTGTCATCGTGTATTTACTT 1077
1132 GTACACCAATGATTTTACATCTGCTTACCTGCTTCCAGGTAATTCGAATTA 1137
1078 GTACACCAATGATTTTACATCTGCTTACCTGCTTCCAGGTAATTCGAATTA 1137
1192 ATTTGACATCCGCTAAGAGGCGCATCTCTCAACCTTCTCCAGTCAATTA 1251
1138 ATTTGACATCCGCTAAGAGGCGCATCTCTCAACCTTCTCCAGTCAATTA 1197
1252 GCAAAATATTATAGAGCCCTTACGTTGAGCAATCAATTTGATGTAATTAAGAAAAT 1311
1198 GCAAAATATTATAGAGCCCTTACGTTGAGCAATCAATTTGATGTAATTAAGAAAAT 1257
1312 AGATTAATCCCTTATCAGTAATGCTTACGAGCAAACTAGTAATCATTAGAGTAT 1371
1258 AGATTAATCCCTTATCAGTAATGCTTACGAGCAAACTAGTAATCATTAGAGTAT 1317
1372 GCGCTATGTTTGTGTTAGGTTGTTATCAATCAATTTTACACATTCGATCA 1431
1318 GCGCTATGTTTGTGTTAGGTTGTTATCAATCAATTTTACACATTCGATCA 1377
1432 ATGTAAATTAAGAACCAATATACATCAAGTAAGTAATGTGTATCTGCAT 1491
1378 ATGTAAATTAAGAACCAATATACATCAAGTAAGTAATGTGTATCTGCAT 1437
1492 TAAAGTATCCAGTATTTGATCACTTATTAATAAAGAAATGATTTATCGTA 1551
1438 TAAAGTATCCAGTATTTGATCACTTATTAATAAAGAAATGATTTATCGTA 1497
1552 ATAACTGGTTTATGTCAGTACTGTAAATATACTAGAGTTAAATTAATGTTTACTC 1611
1498 ATAACTGGTTTATGTCAGTACTGTAAATATACTAGAGTTAAATTAATGTTTACTC 1557


```

QY      1612 TGCCTCACCAACACATGCTAGGATTAACCCCAAAATTAATATTAACTTGATTCGATTG 1671
      |||
Db      1558 TGCCTCACCAACACATGCTAGGATTAACCCCAAAATTAATATTAACTTGATTCGATTG 1617
QY      1672 GTATTAAGAGAGACTGGCTGCTATTAATTAGATTAATTGAGGACAGAGAGCTCTTATC 1731
      |||
Db      1618 GTATTAAGAGAGACTGGCTGCTATTAATTAGATTAATTGAGGACAGAGAGCTCTTATC 1677
QY      1732 CTAATCTGATTTAGTATGTTCTGTATTAATGAGAAATGTTACCAAAATTAATCTTTAGTG 1791
      |||
Db      1678 CTAATCTGATTTAGTATGTTCTGTATTAATGAGAAATGTTACCAAAATTAATCTTTAGTG 1737
QY      1792 ATTACATGTCATCTTTTATAGGGGACATGTTCTGTATTAAGGATTAATAATTAATTA 1851
      |||
Db      1738 ATTACATGTCATCTTTTATAGGGGACATGTTCTGTATTAAGGATTAATAATTAATTAATTA 1797
QY      1852 GTA 1854
      |||
Db      1798 GTA 1800

RESULT 7
AC090439/c 199002 bp DNA linear HTG-29-JUN-2002
LOCUS Homo sapiens chromosome 12 clone RP11-418117, WORKING DRAFT
DEFINITION AC090439.8 GI:21629092
ACCESSION AC090439
VERSION HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
KEYWORDS Homo sapiens (human)
SOURCE Homo sapiens
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
1 (bases 1 to 199002)
Mzany,D.M., Adams,C., Adio-Osula,B., Ali-osman,F.R., Allen,C.,
Alshrocks,S.L., Amaralunga,H.C., Are,J.R., Ayale,M., Baks,F.,
Barbata,J., Benton,J., Bimaga,K., Blankenburg,K., Bonnin,D.,
Bouch,J., Bowie,S., Brieva,M., Brown,M., Brown,M., Bryant,N.P.,
Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
Carroll,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
Chen,G., Chen,R., Chen,Z., Chowdhury,I., Christopoulos,C.,
Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
Devila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
Doutwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Franz,P.,
Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
Gorrell,J.H., Guevara,M., Gunaratne,P., Hale,S., Hamilton,K.,
Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J.,
Hernandez,O., Hodgson,A., Hognes,M., Holloway,C., Hollins,B.,
Homs,F., Howard,S., Huber,J., Huijck,S., Hume,J., Jackson,L.E.,
Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,
Karleson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,
Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,
Li,J., Li,Z., Lichtenage,O., Lien,C., Liu,J., Liu,W., Louleaged,H.,
Lozawo,R.J., Lu,X., Lucier,A., Lucier,C., Luna,R., Ma,J.,
Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E.,
Massey,E., Mawhinney,E., McLeod,M.P., Meador,M., Mei,G., Metzger,M.,
Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,
Moser,M., Neal,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N.,
Nguyen,N., Nickerson,E., Nwokemkwo,S., Ogulu,M., Okwundu,G.,
Ogunyeye,N., Oyiedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,
Peterson,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y.,
Rivers,M., Rojas,A., Rojupokan,I., Rolfe,M., Ruiz,S., Savery,G.,
Scherer,S., Scott,G., Shen,H., Shoshitari,N., Sisson,I.,
Sodergren,E., Somaite,T., Sparks,A., Stanley,H., Stone,I.,
Sutton,A., Svatek,A., Taber,P., Tamerisa,A., Tamerisa,K., Tang,H.,
Tansley,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,O.,
Usmani,K., Vasquez,L., Vera,V., Villalón,D., Vinson,R., Wang,Q.,
Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S.,
Williams,G., Williamson,A., Wlaczek,R., Woodden,S., Worley,K.,
Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorilla,S., Nelson,D.,

```

```

TITLE
JOURNAL
REFERENCE
AUTHORS
Worley,K.C.
Direct Submission
Submitted (22-FEB-2001) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 199002)
Worley,K.C.
Direct Submission
Submitted (29-JUN-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Jun 29, 2002 this sequence version replaced gi:20279190.

COMMENT
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Drafting Center Code: WIKR
Contact: hgsc-help@bcm.tmc.edu
Center project name: HDAC
Center clone name: RP11-418117
----- Project Information
----- Summary Statistics
Sequencing vector: M13;
Chemistry: Dye-Primer Body: 65% of reads
Chemistry: Dye-terminator Big Dye: 35% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 195705 bases at least Q40
Consensus quality: 197756 bases at least Q30
Consensus quality: 199521 bases at least Q20
Estimated insert size: 211863; sum-of-contigs estimation
Quality coverage: 9x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 14 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1
2051: contig of 2051 bp in length
2052
2152: gap of unknown length
2152
4816: contig of 2665 bp in length
4817
4916: gap of unknown length
4917
8256: contig of 3340 bp in length
8257
8356: gap of unknown length
8357
11062: contig of 2706 bp in length
11063
11162: gap of unknown length
11163
14224: contig of 3062 bp in length
14224
14324: gap of unknown length
14325
18185: contig of 3861 bp in length
18185
18285: gap of unknown length
18286
23102: contig of 4817 bp in length
23102
23202: gap of unknown length
23203
28667: contig of 5465 bp in length
28667
28767: gap of unknown length
28768
34634: contig of 5867 bp in length
34634
34734: gap of unknown length
34735
40230: contig of 5456 bp in length
40231
40330: gap of unknown length
40331
46114: contig of 5784 bp in length
46114
46214: gap of unknown length
46215
53091: contig of 6877 bp in length
53091
53191: gap of unknown length
53192
125180: contig of 71989 bp in length
125181
199002: contig of 73722 bp in length.

```

FEATURES
Source Location/Qualifiers
1..19902
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="12"
/clone="RP11-418117"

ORIGIN

Query Match 76.9%; Score 1439; DB 2; Length 19902;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 1799; Conservative 0; Mismatches 0; Indels 4; Gaps 2;

```
QY 53 GCTGAGCCCGTTAGTGCAGCCGCGGAGACACGCGCCGCAATGCCGCTACTCGGTC 112
Db 22174 GCTGAGCCCGTTAGTGCAGCCGCGGAGACACGCGCCGCAATGCCGCTACTCGGTC 22115
QY 113 CCCCCACACGCTCTGTTCGTCAAGGACGTCGCGGACGACACGAGCTGAGAGCTTGC 172
Db 22114 CCCCCACACGCTCTGTTCGTCAAGGACGTCGCGGACGACACGAGCTGAGAGCTTGC 22055
QY 173 GGCCTGATTTGTCGTATGAGTCCCTATAGTGTATGTTCACCTGATTTCTACA 232
Db 22054 GGCCTGATTTGTCGTATGAGTCCCTATAGTGTATGTTCACCTGATTTCTACA 21995
QY 233 CTCGCGCTCCAAAGAGATTGCTTATGTTCAATTTGAGAGATGTCGTATGCTGAAGCG 292
Db 21994 CTCGCGCTCCAAAGAGATTGCTTATGTTCAATTTGAGAGATGTCGTATGCTGAAGCG 21935
QY 233 CTTTACATTAATTGGACAGAAAGTGGATTGTGACGCGGACATGAAATACAGTTTGC 352
Db 21934 CTTTACATTAATTGGACAGAAAGTGGATTGTGACGCGGACATGAAATACAGTTTGC 21875
QY 353 AGGGGGATCGAAAGACACCAATCAGATGAAAGCCAGGAAAGAGAAATGTACAGTT 412
Db 21874 AGGGGGATCGAAAGACACCAATCAGATGAAAGCCAGGAAAGAGAAATGTACAGTT 21815
QY 413 CTTACGCTATGATATTTATGACAGATACAGCGTTCTTAAAGCCGAAAGTTTGAAGA 472
Db 21814 CTTACGCTATGATATTTATGACAGATACAGCGTTCTTAAAGCCGAAAGTTTGAAGA 21755
QY 473 GGAGAGTCAAGAGTGGTCTTTGATTACATCATTAAGAGATGATAGTCTTAAAGCA 532
Db 21754 GGAGAGTCAAGAGTGGTCTTTGATTACATCATTAAGAGATGATAGTCTTAAAGCA 21697
QY 533 GTAGACCGCATGGAAGACCCAGCGCGTAG-AGAAGCCATTCGACAAATGATAGACCAACT 591
Db 21696 -TAGACCGCATGGAAGACCCAGCGCGTAG-AGAAGCCATTCGACAAATGATAGACCAACT 21638
QY 592 GCAGCTGGAATACCCAGTACAGTTCGTCTTACTACCTTCAGAAAGATCTGAAGCCGA 651
Db 21637 GCAGCTGGAATACCCAGTACAGTTCGTCTTACTACCTTCAGAAAGATCTGAAGCCGA 21578
QY 652 AAAAGAACCAAGAAAGAGGAGTTCAGAGCACCAAGGAGTGGTGAAGGTGCTGCAGTAT 711
Db 21577 AAAAGAACCAAGAAAGAGGAGTTCAGAGCACCAAGGAGTGGTGAAGGTGCTGCAGTAT 21518
QY 712 GAATACGTACGAATATTTGACTGCTGTGAAAAAGATPAAAGATGTTATGAAAAC 771
Db 21517 GAATACGTACGAATATTTGACTGCTGTGAAAAAGATPAAAGATGTTATGAAAAC 21458
QY 772 ACATGGAATATTTGAAGTCCCTTCAGTTTGAAGTGAAGTATTTGGAACAATAAAG 831
Db 21457 ACATGGAATATTTGAAGTCCCTTCAGTTTGAAGTGAAGTATTTGGAACAATAAAG 21398
QY 832 AAATTAACCTTTGACTGTGGAACCTAATCCTTAATAATGAAATAGGTTATATTGATTC 891
Db 21397 AAATTAACCTTTGACTGTGGAACCTAATCCTTAATAATGAAATAGGTTATATTGATTC 21338
QY 892 ATGGGTAACAGGTCACTAATTAATTTATTTGAACCTAGAGTGTGAAATCAAGGAAGAC 951
Db 21337 ATGGGTAACAGGTCACTAATTAATTTATTTGAACCTAGAGTGTGAAATCAAGGAAGAC 21278
```

```
QY 952 AGCCATAGTCTCTTACAGAGTGCCTCTGTGTCGTCTCAAACTGAATTTGGTGGGAAAG 1011
Db 21277 AGCCATAGTCTCTTACAGAGTGCCTCTGTGTCGTCTCAAACTGAATTTGGTGGGAAAG 21218
QY 1012 GTATGGTCCAAATPAAAGTTCCATTTTCCATTTATGGCAATCTTGGCTTTGTTAT 1071
Db 21217 GTATGGTCCAAATPAAAGTTCCATTTTCCATTTATGGCAATCTTGGCTTTGTTAT 21158
QY 1072 TTTGGTCCAGTGTCTTCTGCTTAATCATTTGCTTTGTCGATCTGTTATTAATCT 1131
Db 21157 TTTGGTCCAGTGTCTTCTGCTTAATCATTTGCTTTGTCGATCTGTTATTAATCT 21098
QY 1132 GTACACCAACATGACGTTTACATCTGCTTAACTACCTCCCTCCAGGTAATTTCAATTA 1191
Db 21097 GTACACCAACATGACGTTTACATCTGCTTAACTACCTCCCTCCAGGTAATTTCAATTA 21038
QY 1192 ATTTGACATCCAGTAAAGAGGCCCATCTCTTCCACTCTTCTCCATGCTGATATCA 1251
Db 21037 ATTTGACATCCAGTAAAGAGGCCCATCTCTTCCACTCTTCTCCATGCTGATATCA 20978
QY 1252 GCAAATATTTATTTGAGCCCTTACTGTGGGCAATCATTTGATGATTAATGAGAAAAAT 1311
Db 20977 GCAAATATTTATTTGAGCCCTTACTGTGGGCAATCATTTGATGATTAATGAGAAAAAT 20918
QY 1312 AGATTAATTCCTTATTCAGTAATGTCTACTGACACAACTTGAATCATTTACAGTAT 1371
Db 20917 AGATTAATTCCTTATTCAGTAATGTCTACTGACACAACTTGAATCATTTACAGTAT 20858
QY 1372 GGCTCATTTGTTTGTGTTGAGGTGTATTCATTAACAATTTTACCCAGTGGTATCA 1431
Db 20857 GGCTCATTTGTTTGTGTTGAGGTGTATTCATTAACAATTTTACCCAGTGGTATCA 20798
QY 1432 ATGTAATTAATGAACACAAATPACATGATCAAGATTAATTTGTTGTTATTCGCCATT 1491
Db 20797 ATGTAATTAATGAACACAAATPACATGATCAAGATTAATTTGTTGTTATTCGCCATT 20738
QY 1492 TAAAGTATCCAGTATTTGATCACTTATTAATTAATTAATGAATAATGATTTAATCTGTA 1551
Db 20737 TAAAGTATCCAGTATTTGATCACTTATTAATTAATTAATGAATAATGATTTAATCTGTA 20678
QY 1552 ATAAACTGGTTATTTGTCAGTGAAGTGAATTAATCTAGATTAATTAATGTTTACTC 1611
Db 20677 ATAAACTGGTTATTTGTCAGTGAAGTGAATTAATCTAGATTAATTAATGTTTACTC 20618
QY 1612 TGCCCTCAACCAACACATGCTAGATATTAACCCCAAAATTAATTAATTTAATCTGATTAG 1671
Db 20617 TGCCCTCAACCAACACATGCTAGATATTAACCCCAAAATTAATTAATTTAATCTGATTAG 20558
QY 1672 GTATTAAGAGAGACTGGGTGCTATTAATTAATTAATTTTGAAGGACAGAGAGCTGTATC 1731
Db 20557 GTATTAAGAGAGACTGGGTGCTATTAATTAATTAATTTTGAAGGACAGAGAGCTGTATC 20498
QY 1732 CTAACTGATTTAGTATGTTCTGTAATTAATTAATTAATTTTGAAGGATTAATTAATG 1791
Db 20497 CTAACTGATTTAGTATGTTCTGTAATTAATTAATTAATTTTGAAGGATTAATTAATG 20438
QY 1792 ATTACATGATATTTATTAAGGAGCATGTTCTGTGTATTAAGGAATTAATTAATTTA 1851
Db 20437 ATTACATGATATTTATTAAGGAGCATGTTCTGTGTATTAAGGAATTAATTAATTTA 20378
QY 1852 GTA 1854
Db 20377 GTA 20375
```

RESULT 8
AY048592 16745 bp DNA linear PRI 26-SEP-2001
LOCUS AY048592
DEFINITION Homo sapiens TUS-associated SR proteins (TUSP) gene, complete cds,
alternatively spliced.
ACCESSION AY048592
VERSION AY048592.1 GI:15787481
KEYWORDS

	polyA_site	/note="for TAsR-2 transcript" 14879 /gene="TASR"
	polyA_site	/note="alternative site for TASR-1 transcript" 16515 /gene="TASR" /note="for TASR-1 transcript"
	ORIGIN	
	Query Match	68.2%; Score 1276; DB 9; Length 16745;
	Best Local Similarity	100.0%; Pred. No. 0;
	Matches 1276; Conservative	0; Mismatches 0; Indels 0; Gaps 0;
QY	582 AGACCAAACTGACGTGGATATACCAGTAAGTCGTCTTGACTTACTACAATTCCAGAAGATC	641
DB	13603 AGACCAAACTGACGTGGATATACCAGTAAGTCGTCTTGACTTACTACAATTCCAGAAGATC	13662
QY	642 TGAAAAGCGGAAAAAGAACCCAAAGGAAGGCAGCTTAGCCGCCAAAGGGGTGGGAAAGTT	701
DB	13663 TGAAAAGCGGAAAAAGAACCCAAAGGAAGGCAGCTTAGCCGCCAAAGGGGTGGGAAAGTT	13722
QY	702 GCTGCAATGATATCTGTAAGTAATTTTTGACTCTGCTGTCGAAAAAGATPAAAGATGTT	761
DB	13723 GCTGCAATGATATCTGTAAGTAATTTTTGACTCTGCTGTCGAAAAAGATPAAAGATGTT	13782
QY	762 ATCGAAAACTACATGAGAAATATTTGAAGTCCCTTCAAATTGAAAGTAAGCATTTTAGCAC	821
DB	13783 ATCGAAAACTACATGAGAAATATTTGAAGTCCCTTCAAATTGAAAGTAAGCATTTTAGCAC	13842
QY	822 AATPAAAAAGAAATTCGAATCTTGACTGTGGGAAACTAATCCCTTAATATGAATAGCTTT	881
DB	13843 AATPAAAAAGAAATTCGAATCTTGACTGTGGGAAACTAATCCCTTAATATGAATAGCTTT	13902
QY	882 ATATTGATTCATGGGTAAACAGTCCATAATPAATTTATGGAACCTAGATGTCGATAT	941
DB	13903 ATATTGATTCATGGGTAAACAGTCCATAATPAATTTATGGAACCTAGATGTCGATAT	13962
QY	942 CAAGGAAGACGCCATAGTCTCTTACAGTGCCTCTGTGGTCTGCTCAAACTGAATTGG	1001
DB	13963 CAAGGAAGACGCCATAGTCTCTTACAGTGCCTCTGTGGTCTGCTCAAACTGAATTGG	14022
QY	1002 GTGGGAAAAGGTAATGCTCCATATTAAGTTCATTTTGGCATTTATGGCAAACTCTTGC	1061
DB	14023 GTGGGAAAAGGTAATGCTCCATATTAAGTTCATTTTGGCATTTATGGCAAACTCTTGC	14082
QY	1062 CTTTGTTAATTTTGTGCCAGTGTCTTTCGTCTTAATCATTTGCTTTGTGGATCTGTGT	1121
DB	14083 CTTTGTTAATTTTGTGCCAGTGTCTTTCGTCTTAATCATTTGCTTTGTGGATCTGTGT	14142
QY	1122 TTATTATCTTGAACCCACATGCAATTCATCTGTCTTAACTAGTCTCTCCAGGTAAA	1181
DB	14143 TTATTATCTTGAACCCACATGCAATTCATCTGTCTTAACTAGTCTCTCCAGGTAAA	14202
QY	1182 TTCCAATTAATTTGACATTCACGCTAAGAGGCCCATCTCTTCTCACTCTTCTAGTC	1241
DB	14203 TTCCAATTAATTTGACATTCACGCTAAGAGGCCCATCTCTTCTCACTCTTCTAGTC	14262
QY	1242 AGTATATTCAGCAAAATTTATTTGAGCCCTTACCTGAGGGCAAAATCATTTGTAAGTAT	1301
DB	14263 AGTATATTCAGCAAAATTTATTTGAGCCCTTACCTGAGGGCAAAATCATTTGTAAGTAT	14322
QY	1302 TGAAAAAATATGATTAATCTCTTATTCAGTAATGTCTACTAGACACAATCTAGTGAATC	1361
DB	14323 TGAAAAAATATGATTAATCTCTTATTCAGTAATGTCTACTAGACACAATCTAGTGAATC	14382
QY	1362 ATTACGTAATGCGCTCATATGTTTGTGTAAGGTGTATTCATTAACAATATTTTACAC	1421
DB	14383 ATTACGTAATGCGCTCATATGTTTGTGTAAGGTGTATTCATTAACAATATTTTACAC	14442
QY	1422 ATTTCGATCAATGTATTTATAGACACAATATAGCATCAAGGATPAAAGTAATGTGTGTT	1481
DB	14443 ATTTCGATCAATGTATTTATAGACACAATATAGCATCAAGGATPAAAGTAATGTGTGTT	14502


```

QY 1362 ATTACAGTATGGCTCATGTTTGTGTAGAGTGTATTCTATTAACAATATTATACACC 1421
DB 36771 ATTACAGTATGGCTCATGTTTGTGTAGAGTGTATTCTATTAACAATATTATACACC 36712
QY 1422 ATTCGATCAATGATATTATAGAACACATATACGATCAAGGATAGTATTTGTGTGTT 1481
DB 36711 ATTCGATCAATGATATTATAGAACACATATACGATCAAGGATAGTATTTGTGTGTT 36652
QY 1482 ATCTGCATTAAAGATATCCAGTATTGATCACTTATTAATAATGAAGAAATGAT 1541
DB 36651 ATCTGCATTAAAGATATCCAGTATTGATCACTTATTAATAATGAAGAAATGAT 36592
QY 1542 TTAACCTGTAATAAATCGTTTATTTGTGACGAGTGTATATCTAGAGTTATATAA 1601
DB 36591 TTAACCTGTAATAAATCGTTTATTTGTGACGAGTGTATATCTAGAGTTATATAA 36532
QY 1602 TTGTTTACTGTCGCCACCAACACATGCTAGATATACCCCAAAATTAAGTATTTAAC 1661
DB 36531 TTGTTTACTGTCGCCACCAACACATGCTAGATATACCCCAAAATTAAGTATTTAAC 36472
QY 1662 TTTCGATTAGGATTAAGAGAGACTGGTGTATATTAATTAATTTTGGAGACAGAG 1721
DB 36471 TTTCGATTAGGATTAAGAGAGACTGGTGTATATTAATTAATTTTGGAGACAGAG 36412
QY 1722 AGCTGTATCTTACTGATTTAGTATGTTCTGTATTAATGAAGAAATGTTACCAATTATA 1781
DB 36411 AGCTGTATCTTACTGATTTAGTATGTTCTGTATTAATGAAGAAATGTTACCAATTATA 36352
QY 1782 CTTTTTAGTGATTTACATGATCATTTTATAGGGGACATGCTGTGTATAGGAATTAAT 1841
DB 36351 CTTTTTAGTGATTTACATGATCATTTTATAGGGGACATGCTGTGTATAGGAATTAAT 36292
QY 1842 AACTTTATAGTATCA 1857
DB 36291 AACTTTATAGTATCA 36276

```

```

RESULT 10
AL590609/c 172307 bp DNA linear PRI 08-DEC-2003
LOCUS Human DNA sequence from clone Rp11-4M23 on chromosome 1, complete
DEFINITION
AL590609
ACCESSION
AL590609.15 GI:39573520
VERSION
HTG.
KEYWORDS
SOURCE
Homo sapiens (human)
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
1 (bases 1 to 172307)
AUTHORS
Wallis, J.
TITLE
Direct Submission
JOURNAL
Submitted (08-DEC-2003) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
humquerry@sanger.ac.uk; Clone requests: clonerequest@sanger.ac.uk
On Dec 8, 2003 this sequence version replaced gi:15795484.
----- Genome Center
Center: Wellcome Trust Sanger Institute
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquerry@sanger.ac.uk

```

During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above. This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such

as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest, except on the rare occasion of the clone being a YAC. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Emi, EMBL; Sw, SWISSPROT; Tr, TREMBL; Wp, WORMEP; Information on the WORMEP database can be found at http://www.sanger.ac.uk/Projects/C_elegans/wormep This sequence was generated from part of bacterial clone contigs of human chromosome 1, constructed by the Sanger Centre Chromosome 1 Mapping Group. Further information can be found at <http://www.sanger.ac.uk/HGP/Chr1> Rp11-4M23 is from the library RPCI-11.1 constructed by the group of Pieter de Jong. For further details see <http://www.chori.org/bacpac/home.htm> VECTOR: pBAC3.6.

FEATURES

```

source
1..172307
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="1"
/clone="Rp11-4M23"
/clone_id="RPCI-11.1"

```

ORIGIN

Query Match 68.2%; Score 1276; DB 9; Length 172307;

Best Local Similarity 100.0%; Pred. No. 0; Matches 1276; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 582 AGACCAACAGTCAGCTGGAATATCCAGTACAGTCTGCTTACTACATCTCAAGAAAGTC 641
DB 100747 AGACCAACAGTCAGCTGGAATATCCAGTACAGTCTGCTTACTACATCTCAAGAAAGTC 100688
QY 642 TGAAGGCGAAAAAGAACCAAGAGGCGAGTTCAAGCAGCAAGGAGGAGGAGGAGT 701
DB 100687 TGAAGGCGAAAAAGAACCAAGAGGCGAGTTCAAGCAGCAAGGAGGAGGAGT 100628
QY 702 GCTGCAGTATGATATCTGTACGAATATTTTGTACTCTGCTGTGAAAGATTAAGATGT 761
DB 100627 GCTGCAGTATGATATCTGTACGAATATTTTGTACTCTGCTGTGAAAGATTAAGATGT 100568
QY 762 ATCGAAATACATCATGGAATATGAAGTCCCTCAAGTTGAAGTAAAGATTTTAAGAC 821
DB 100567 ATCGAAATACATCATGGAATATGAAGTCCCTCAAGTTGAAGTAAAGATTTTAAGAC 100508
QY 822 AAATTAAGGAAATTCACCTTTGTACTTGTGAAACTAATCCCTAATATGAATAGGTT 881
DB 100507 AAATTAAGGAAATTCACCTTTGTACTTGTGAAACTAATCCCTAATATGAATAGGTT 100448
QY 882 ATATTTGATATGAGGATACAGGTCATTAATAATTTGAAACTAGAGATGCTGAATAT 941
DB 100447 ATATTTGATATGAGGATACAGGTCATTAATAATTTGAAACTAGAGATGCTGAATAT 100388
QY 942 CAAGAAAGACAGCATAGTCTCTTACAGTCCCTGTGTGCTGCTCAAGATGAATTGG 1001
DB 100387 CAAGAAAGACAGCATAGTCTCTTACAGTCCCTGTGTGCTGCTCAAGATGAATTGG 100328
QY 1002 GTGGGAAAAGGATATGATTAATAAGTTCCATTTTGGCATTAATTTGGCAATCTTGC 1061
DB 100327 GTGGGAAAAGGATATGATTAATAAGTTCCATTTTGGCATTAATTTGGCAATCTTGC 100268
QY 1062 CTTTGTATTTTGTGGCCAGTGTCTTCAATCAATTTGCTTTGTGGCATCTGTGT 1121
DB 100267 CTTTGTATTTTGTGGCCAGTGTCTTCAATCAATTTGCTTTGTGGCATCTGTGT 100208
QY 1122 TTATTTACTGTACACCATGAGTTTACATCTGTACACTGCTTCCAGGATTA 1181
DB 100207 TTATTTACTGTACACCATGAGTTTACATCTGTACACTGCTTCCAGGATTA 100148
QY 1182 TTTCATTTATTTGACATCCAGTAAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1241

```

```

Db 100147 TTCGAATTATTTGACATCCAGCTAAGAGGCCCATCTCTTCACCTCTTCTAGTC 100088
QY 1242 AGTATATTCAGCAAAATATTTATTTAGACCTTACTGTGGGCAATCTGTAATG 1301
Db 100087 AGTATATTCAGCAAAATATTTATTTAGACCTTACTGTGGGCAATCTGTAATG 100028
QY 1302 TGAGAAAAATGATATTTCCCTTATTTAGCAATATGCTACAGACCAATCTAGATC 1361
Db 100027 TGAGAAAAATGATATTTCCCTTATTTAGCAATATGCTACAGACCAATCTAGATC 99968
QY 1362 ATTACAGATGAGCTCATTTGTTTGTGAGAGTGTATTCATACAAATTTTACACC 1421
Db 99967 ATTACAGATGAGCTCATTTGTTTGTGAGAGTGTATTCATACAAATTTTACACC 99908
QY 1422 ATTGCAATCAATGATATTTAGCAACAATATAGCATACAGATATGATTTGTGTT 1481
Db 99907 ATTGCAATCAATGATATTTAGCAACAATATAGCATACAGATATGATTTGTGTT 99848
QY 1482 ATCTGCCATTTAAAGATATCCAGATTTGATCACTTTTAAATTAATGAAAAATGAT 1541
Db 99847 ATCTGCCATTTAAAGATATCCAGATTTGATCACTTTTAAATTAATGAAAAATGAT 99788
QY 1542 TTAATCTGTAATTAATCTGTTTATTTGTCAGTCACTGTAATATCTAGATTAATATA 1601
Db 99787 TTAATCTGTAATTAATCTGTTTATTTGTCAGTCACTGTAATATCTAGATTAATATA 99728
QY 1602 TTGTTTACTCTGCTCCACCAACACATGCTAGGATATTAATCCCAATTAATTTAAC 1661
Db 99727 TTGTTTACTCTGCTCCACCAACACATGCTAGGATATTAATCCCAATTAATTTAAC 99668
QY 1662 TTTCGATTAAGATTAAGAGAGCTGGTGTCTAATTAATTAATTTAGGAGACAGAG 1721
Db 99667 TTTCGATTAAGATTAAGAGAGCTGGTGTCTAATTAATTAATTTAGGAGACAGAG 99608
QY 1722 AGCTGTATTCCTAAGTATTAATGTTCTGTGTAATTTGAGAAATGTTTACCAATATA 1781
Db 99607 AGCTGTATTCCTAAGTATTAATGTTCTGTGTAATTTGAGAAATGTTTACCAATATA 99548
QY 1782 CTTTATGATTTACATGATATTTTATAGGGACATGTTCTGTGTAATTTAGGATTAAT 1841
Db 99547 CTTTATGATTTACATGATATTTTATAGGGACATGTTCTGTGTAATTTAGGATTAAT 99488
QY 1842 AACTTTATAGTATCA 1857
Db 99487 AACTTTATAGTATCA 99472

```

```

RESULT 11
HSM801468 2073 bp mRNA linear pri 18-FEB-2000
LOCUS HSM801468 2073 bp mRNA linear pri 18-FEB-2000
DEFINITION Homo sapiens mRNA, cDNA DKFZp434N1717 (from clone DKFZp434N1717).
ACCESSION AL133655
VERSION AL133655.1 GI:6599245
KEYWORDS
SOURCE
ORGANISM Homo sapiens (human)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (bases 1 to 2073)
AUTHORS Bloeker,H., Boecker,M., Brandt,P., Mewes,H.W., Gaassenhuber,J. and
Wiemann,S.
TITLE Direct Submission
JOURNAL Submitted (15-DEC-1999) MIPS, Am Klopferpitz 18a, D-82152
Martinsried, GERMANY
COMMENT
Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
Research Center (DKFZ); Email:s.wiemann@dkfz-heidelberg.de;
Sequenced by GPF (National Research Centre for Biotechnology Ltd.,
Braunschweig/Germany) within the cDNA sequencing consortium of the
German Genome Project.
This clone (DKFZp434N1717) is available at the RZPD in Berlin.
Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059
Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de Further
information about the clone and the sequencing project is available

```

```

FEATURES
source at http://www.mips.biochem.mpg.de/proj/cDNA/.
location/Qualifiers
1..2073
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/map="716_F_2; 850_F_7; 882_B_6; 900_C_1; 912_B_9;
964_H_2; 800_D_12; 809_H_11; (934,936)_C_10"
/clone="DKFZp434N1717"
/tissue_type="testis"
/clone_id="434 (synonym: htes3). Vector pSport1; host
DH10B; sites NotI + SalI"
/dev_stage="adult"
2001..2006
polya_signal
polya_site
2016
ORIGIN

```

```

Query Match 68.0%; Score 1273; DB 9; Length 2073;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1273; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 582 AGACCAATCTGACCTGGAATATCCAGATGTTCTGCTACTACATTCAGAAAGATC 641
Db 748 AGACCAATCTGACCTGGAATATCCAGATGTTCTGCTACTACATTCAGAAAGATC 807
QY 642 TGAAAGCGGAAAAAGAACCAAGAGGAGGAGTCAAGGACCAAGGGGTGGAGAGT 701
Db 808 TGAAAGCGGAAAAAGAACCAAGAGGAGGAGTCAAGGACCAAGGGGTGGAGAGT 867
QY 702 GCTGAGATGATATCTGATGATATTTTGAATCTGCTGTAATTAATTAATTAAT 761
Db 868 GCTGAGATGATATCTGATGATATTTTGAATCTGCTGTAATTAATTAATTAATTAAT 927
QY 762 ATCCGAAATCTCATGGAATTAATTTGAAGTCCCTCAAGTTGAAGTAAGATTTTGAAG 821
Db 928 ATCCGAAATCTCATGGAATTAATTTGAAGTCCCTCAAGTTGAAGTAAGATTTTGAAG 987
QY 822 AAATTAAGGAAATTCATTGTTGTTGGAATCTAATCCCTTAATTAATTAATTAAT 881
Db 988 AAATTAAGGAAATTCATTGTTGTTGGAATCTAATCCCTTAATTAATTAATTAATTAAT 1047
QY 882 ATATTGATTCATGGGTAACTAGGTCAATTAATTAATTTGGAATCTAGATAT 941
Db 1048 ATATTGATTCATGGGTAACTAGGTCAATTAATTAATTTGGAATCTAGATAT 1107
QY 942 CAAGGAAGACAGCCATGATCTCTTACAGGCTCTGTTGCTGTCTCAATCTGAATGG 1001
Db 1108 CAAGGAAGACAGCCATGATCTCTTACAGGCTCTGTTGCTGTCTCAATCTGAATGG 1167
QY 1002 GTGGGAAAAAGTATGATGTCATATAAAAGTTTCATTTTGCATTTATGCAATCTTGC 1061
Db 1168 GTGGGAAAAAGTATGATGTCATATAAAAGTTTCATTTTGCATTTATGCAATCTTGC 1227
QY 1062 CTTTGTATTTTGTGTGACAGTGTTCCTGCTTAATCAATTTGTTGGCATCTGTGT 1121
Db 1228 CTTTGTATTTTGTGTGACAGTGTTCCTGCTTAATCAATTTGTTGGCATCTGTGT 1287
QY 1122 TTATTACTTGTACACCATGATGATTTATCTGCTTAATCAATCTTCCAGGTAA 1181
Db 1288 TTATTACTTGTACACCATGATGATTTATCTGCTTAATCAATCTTCCAGGTAA 1347
QY 1182 TTCGAATTAATTTGACATCCAGCTAAGAGGCCCATCTCTTCTCACTTTCTTAGTC 1241
Db 1348 TTCGAATTAATTTGACATCCAGCTAAGAGGCCCATCTCTTCTCACTTTCTTAGTC 1407
QY 1242 AGTATATTCAGCAAAATATTTATTTAGACCTTACTGTGGGCAATCTGTAATG 1301
Db 1408 AGTATATTCAGCAAAATATTTATTTAGACCTTACTGTGGGCAATCTGTAATG 1467
QY 1302 TGAGAAAAATGATATTTCCCTTATTTAGCAATATGCTACAGACCAATCTAGATC 1361
Db 1468 TGAGAAAAATGATATTTCCCTTATTTAGCAATATGCTACAGACCAATCTAGATC 1527

```

QY 1362 ATTACAGTANGGCGCTCATTTGTTTGTGAGGTGTATTCATATCAATATTTTACACC 1421
DB 1528 ATTACAGTATGCGCTCATTTGTTTGTGAGGTGTATTCATATCAATATTTTACACC 1587
QY 1422 ATTGCGATCAATGTATTTATAGAACACATATATCATGATCAAGGATAGTATTTGCGGTT 1481
DB 1588 ATTGCGATCAATGTATTTATAGAACACATATATCATGATCAAGGATAGTATTTGCGGTT 1647
QY 1482 ATCTGCCATTTAAAGTATCCAGTATTTGATCAATTTATTAATGAAAAATGAT 1541
DB 1648 ATCTGCCATTTAAAGTATCCAGTATTTGATCAATTTATTAATGAAAAATGAT 1707
QY 1542 TTAATCTGTAAATACTGGTTTATTTGTGAGGACCTGTATTAATCTAGATTTATATAA 1601
DB 1708 TTAATCTGTAAATACTGGTTTATTTGTGAGGACCTGTATTAATCTAGATTTATATAA 1767
QY 1602 TTGTTTACTCTGCGCTCACCACCAACATGCTAGATATTAACCCCAAAATTAAGTATTAC 1661
DB 1768 TTGTTTACTCTGCGCTCACCACCAACATGCTAGATATTAACCCCAAAATTAAGTATTAC 1827
QY 1662 TTTGATTTAGGTATTAAGGAGACTGGGTCTATTAATTAATTTTGAAGGACAGAG 1721
DB 1828 TTTGATTTAGGTATTAAGGAGACTGGGTCTATTAATTAATTTTGAAGGACAGAG 1887
QY 1722 AGCTGTATCCCTAACTGATTTAGTATGTTCTGTCTAATTTGAGAAATGTTTCACTTAATA 1781
DB 1888 AGCTGTATCCCTAACTGATTTAGTATGTTCTGTCTAATTTGAGAAATGTTTCACTTAATA 1947
QY 1782 CTTTGTAGTATTTAATCATGATCAATTTTATAGGGGACATGTTCTGTGTATAGGAATAAT 1841
DB 1948 CTTTGTAGTATTTAATCATGATCAATTTTATAGGGGACATGTTCTGTGTATAGGAATAAT 2007
QY 1842 AACTTTTATAGTA 1854
DB 2008 AACTTTTATAGTA 2020

RESULT 12
AK093226 2834 bp mRNA linear PRI 15-JUL-2002
LOCUS Homo sapiens cDNA FLJ35907 f1s, clone TEST12009785, weakly similar
DEFINITION to Mus musculus TLS-associated protein with SR repeats mRNA.
AK093226
ACCESSION AK093226.1 GI:21752034
VERSION Oligo capping, f1s (full insert sequence).
KEYWORDS Homo sapiens (human)
SOURCE Homo sapiens
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE
AUTHORS 1 Iihbashi, T., Kanehori, K., Yosida, M., Watanabe, S., Ishida, S., Ono, Y., Hotuta, T., Hiraoka, S., Murakawa, K., Takiguchi, S., Kusano, J., Watanabe, M., Fujimori, K., Tanai, H., Ishida, M., Yamashita, H., Chiba, Y., Sugiyama, T., Irie, R., Otsuki, T., Wakamatsu, A., Ishii, S., Yamamoto, J., Isono, Y., Kawai, H., Saito, K., Nishikawa, T., Kimura, K., Matsuo, K., Nakamura, Y., Sekine, M., Kikuchi, H., Kanda, K., Magatsuma, M., Takahashi, Fujii, A., Oshima, A., Sugiyama, A., Kawakami, B., Suzuki, Y., Sugano, S., Nagatani, K., Masuno, Y., Nagai, K. and Isogai, T.
TITLE NEDO human cDNA sequencing project
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 2834)
AUTHORS Isogai, T. and Yamamoto, J.
TITLE Direct Submission
JOURNAL Submitted (04-JUL-2002) Takao Isogai, FLJ Project (HRI Team); 2-6-7 Kazusa-Kametani, Kisarazu, Chiba 292-0812, Japan.
COMMENT (E-mail: genomics@hri.co.jp, Tel: 81-438-52-3975, Fax: 81-438-52-3986) NEDO human cDNA sequencing project supported by Ministry of Economy, Trade and Industry of Japan; cDNA full insert sequencing: Research Association for Biotechnology (RAB); cDNA library construction: Helix Research Institute (HRI) (supported by Japan Key Technology Center etc.); 5' - 3' end one pass sequencing: RAB, HRI, and Biotechnology Center, National Institute of Technology and

FEATURES
source
1. 2834
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="TEST12009785"
/tissue_type="testis"
/clone_lib="TEST12"
/note="Cloning vector: pME18Fuj3"

ORIGIN
Query Match 65.4%; Score 1225; DB 9; Length 2834;
Best Local Similarity 99.7%; Pred. No. 0;
Matches 1565; Conservative 0; Mismatches 2; Indels 2; Gaps 2;

Evaluation: clone selection for full insert sequencing: HRI and RAB; annotation: HRI and RAB.
Location/Qualifiers

QY 290 ACGTTTACATTAATTTGACAGAAAGTGGATTTGTGACGGCAGATTGAAATACAGTTTG 349
DB 1266 ACGTTTACATTAATTTGACAGAAAGTGGATTTGTGACGGCAGATTGAAATACAGTTTG 1325
QY 350 CCCAGGGGGATCGAAAGCACCAATTCATGATGAAAGCCAGAGAGAGAAATGTGTACA 409
DB 1326 CCCAGGGGGATCGAAAGCACCAATTCATGATGAAAGCCAGAGAGAGAAATGTGTACA 1385
QY 410 GTTCTTACGCTATGATGATTTATGACAGATACAGATTCAGACGTTCTTGAAGCCGAATTTGAAA 469
DB 1386 GTTCTTACGCTATGATGATTTATGACAGATACAGATTCAGACGTTCTTGAAGCCGAATTTGAAA 1445
QY 470 GGAGAGATCAAGAGTGGTCTTTTGTATTAACAATATAGAGATCGTATAGTCTAGAA 529
DB 1446 GGAGAGATCAAGAGTGGTCTTTTGTATTAACAATATAGAGATCGTATAGTCTAGAA 1505
QY 530 ACAATTAACCGACTCGAAGACACACGCGGTAG-AGAAGCATTCGACATGATAGACCA 588
DB 1506 ACAATTAACCGACTCGAAGACACACGCGGTAGACCAATTCGACATGATAGACCA 1565
QY 589 ACTGACGCTGAATATACCAAGTACGATGCTGCTTATCTTCACTTCAAGAAAGCTTGAAGC 648
DB 1566 ACTGACGCTGAATATACCAAGTACGATGCTGCTTATCTTCACTTCAAGAAAGCTTGAAGC 1625
QY 649 GAAAAAAGAACCAAGAGGACAGTTCAGACCAAGAGGTGGTGAAGGTCTGCGAG 708
DB 1626 GAAAAAAGAACCAAGAGGACAGTTCAGACCAAGAGGTGGTGAAGGTCTGCGAG 1685
QY 709 TATGAATATCTGTACGAATATTTTGAATCTGCTGCTGGAAGATTAAGATTTATCGAAA 768
DB 1686 TATGAATATCTGTACGAATATTTTGAATCTGCTGCTGGAAGATTAAGATTTATCGAAA 1745
QY 769 ACTACATGGAATTAATGAAGTCCCTTCAAGTTTGAAGATTTTATAGCAAAATATA 828
DB 1746 ACTACATGGAATTAATGAAGTCCCTTCAAGTTTGAAGATTTTATAGCAAAATATA 1805
QY 829 AGAAAAATCACTTTGACTTTGTGAAACCTAATCCCTAATATGAATAGAGTTTATATGA 888
DB 1806 AGAAAAATCACTTTGACTTTGTGAAACCTAATCCCTAATATGAATAGAGTTTATATGA 1865
QY 889 TTATATGGGTACAGGTTCATTAATTAATTTATGGAACCTAGATGTCTGAATATCAAGAA 948
DB 1866 TTATATGGGTACAGGTTCATTAATTAATTTATGGAACCTAGATGTCTGAATATCAAGAA 1924
QY 949 GACAGCATAGTCTCTTACAGTGCCTGTGTGTCTGTCTCAAACTGAATTTGGTGGAA 1008
DB 1925 GACAGCATAGTCTCTTACAGTGCCTGTGTGTCTGTCTCAAACTGAATTTGGTGGAA 1984
QY 1009 AAGGATATGATCAATTAATAAGTTCCATTTTGGTCCATTTATGCAATCTTGCTTTGTT 1068
DB 1985 AAGGATATGATCAATTAATAAGTTCCATTTTGGTCCATTTATGCAATCTTGCTTTGTT 2044
QY 1069 TATTTTGGTCCAGTCTTTTGTCTTAAATCATTTGCTTTGTGCGATCTGTGTTATTTTA 1128
DB 2045 TATTTTGGTCCAGTCTTTTGTCTTAAATCATTTGCTTTGTGCGATCTGTGTTATTTTA 2104

Db 609 AAAAAAGCGTCTTTGATTACAACTATAGAGATGATATAGCTTAGAAAACGTAGACCG 668
 Qy 541 ACTGAAAGACCAACGGGGTAG-AGAAAGCAATTCGCAAAATGATAGACCAAACTGACGTGG 599
 Db 669 ACTGAAAGACCAACGGGGTAGAGAAAGCAATTCGCAAAATGATAGACCAAACTGACGTGG 728
 Qy 600 AATACCCGATACAGTTCTGCTTACTTACCTTCAAGAAAGATCTGAAAGCGGAAAAAGAC 659
 Db 729 AATACCCGATACAGTTCTGCTTACTTACCTTCAAGAAAGATCTGAAAGCGGAAAAAGAC 788
 Qy 660 CAAAGAGGGCAGTTCAGACGACCAAGGGGTGGTGAAGGTCGTGCAATAGATCTG 719
 Db 769 CAAAGAGGGCAGTTCAGACGACCAAGGGGTGGTGAAGGTCGTGCAATAGATCTG 848
 Qy 720 TAGGAATAATTTGACTCTGCTGCTGAAAAGATTAAGAAATGTATGCAAACTACATGGA 779
 Db 849 TAGGAATAATTTGACTCTGCTGCTGAAAAGATTAAGAAATGTATGCAAACTACATGGA 908
 Qy 780 TAATGAAAGTCCCTTCAAGTTTGAAGATGAGCAATTTAGACAAATTAAGAAATTTCA 839
 Db 909 TAATGAAAGTCCCTTCAAGTTTGAAGATGAGCAATTTAGACAAATTAAGAAATTTCA 968
 Qy 840 CTTTGACTGTGGAACCTAATCCCTAATATGAAATAGTTTATATGATTCATGGGTAA 899
 Db 969 CTTTGACTGTGGAACCTAATCCCTAATATGAAATAGTTTATATGATTCATGGGTAA 1028
 Qy 900 CAGGTCCATTAATAATTAATGGAACCTAGATGCTGAAATCAAGAAAGACAGCCATAG 959
 Db 1029 CAGGTCCATTAATAATTAATGGAACCTAGATGCTGAAATCAAGAAAGACAGCCATAG 1088
 Qy 960 TCTCTTACAGTCCCTTCTGTGTCTGTCTCAAACTGAAATGGGTGGGAAAAAGTATGCTC 1019
 Db 1089 TCTCTTACAGTCCCTTCTGTGTCTGTCTCAAACTGAAATGGGTGGGAAAAAGTATGCTC 1148
 Qy 1020 CAATATAAAGATTCATTTTGGCAATAT--GGCAATCTGCGCTTGTATATTTGGTG 1078
 Db 1149 CAATATAAAGATTCATTTTGGCAATATTTGGCAATCTGCGCTTGTATATTTGGTG 1208
 Qy 1079 CCAAGTGTCTTCTGCTTATATCAATTTGTTGGCAATCTGTTTATTTATCTGTACAC 1138
 Db 1209 CCAAGTGTCTTCTGCTTATATCAATTTGTTGGCAATCTGTTTATTTATCTGTACAC 1268
 Qy 1139 ACATGCAATTACATCTGCTTACTACTCTCTTCCAGGTAA 1180
 Db 1269 ACATGCAATTACATCTGCTTACTACTCTCTTCCAGGTAA 1310
 RESULT 14
 AX713349 2473 bp DNA linear PAT 15-APR-2003
 LOCUS AX713349
 DEFINITION Sequence 33 from Patent EP1293569.
 ACCESSION AX713349
 VERSION AX713349.1 GI:29888205
 KEYWORDS
 SOURCE
 ORGANISM
 Homo sapiens (human)
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1
 Iosagi,T., Sugiyama,T., Otsuki,T., Wakamatsu,A., Sato,H., Ishii,S.,
 Yamamoto,J.I., Isono,Y., Hio,Y., Otsuka,K., Nagai,K., Irie,R.,
 Tamechika,I., Seki,N., Yoshikawa,T., Otsuka,M., Nagahari,K. and
 Masuko,Y.
 Full-length cDNAs
 Patent: EP 1293569-A 33 19-MAR-2003;
 Helix Research Institute (UP) ; Research Association for
 Biotechnology (UP)
 Location/Qualifiers
 1..2473
 /organism="Homo sapiens"
 /mol_type="unassigned DNA"
 /db_xref="taxon:9606"

ORIGIN
 Query Match 42.0%; Score 786; DB 6; Length 2473;
 Best Local Similarity 99.9%; Pred. No. 0;
 Matches 836; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 Qy 582 AGACCAAACTGACGTGGAATACCCAGTACAGTTCTGCTTACTACATTCAGAAAGATC 641
 Db 1388 AGACCAAACTGACGTGGAATACCCAGTACAGTTCTGCTTACTACATTCAGAAAGATC 1447
 Qy 642 TGAAGCCGAAAAAGAACCAAGAAAGGCGAGTTCAAGCCAGCAAGGAGGTGGTGAAGGT 701
 Db 1448 TGAAGCCGAAAAAGAACCAAGAAAGGCGAGTTCAAGCCAGCAAGGAGGTGGTGAAGGT 1507
 Qy 702 GGTGAGATGAATATCTGTACGAATATTTTGAATCTGGTCTGAAAAGTAAAGATGTT 761
 Db 1508 GGTGAGATGAATATCTGTACGAATATTTTGAATCTGGTCTGAAAAGTAAAGATGTT 1567
 Qy 762 ATCGAAAACATGAGATTAATGAAGTCCCTCAAGTTTGAAGTAAAGATTAAGAC 821
 Db 1568 ATCGAAAACATGAGATTAATGAAGTCCCTCAAGTTTGAAGTAAAGATTAAGAC 1627
 Qy 822 AATTAAGAAATTCATCTTTGATCTTGGAACCTAATCCATAATGATGATGTT 881
 Db 1628 AATTAAGAAATTCATCTTTGATCTTGGAACCTAATCCATAATGATGATGTT 1687
 Qy 882 ATATTGATTCATGGGTAAACAGTCCATTAATAATTTTGAACCTAGATGCTGAATAT 941
 Db 1688 ATATTGATTCATGGGTAAACAGTCCATTAATAATTTTGAACCTAGATGCTGAATAT 1747
 Qy 942 CAAAGAAAGACCCATAGTCTCTTAAGTGGCTCTGTGCTGCTGCTCAAACTGAATTGG 1001
 Db 1748 CAAAGAAAGACCCATAGTCTCTTAAGTGGCTCTGTGCTGCTGCTCAAACTGAATTGG 1807
 Qy 1002 GTGGGAAAAGGATGCTCAATATAAAGTTTCAATTTTGGCAATATGCAAACTTGC 1061
 Db 1808 GTGGGAAAAGGATGCTCAATATAAAGTTTCAATTTTGGCAATATGCAAACTTGC 1867
 Qy 1062 CTTGTTTATTTTGGGCGCAGTGTCTGCTTAATCAATTTGTTGGCATCTGTGT 1121
 Db 1868 CTTGTTTATTTTGGGCGCAGTGTCTGCTTAATCAATTTGTTGGCATCTGTGT 1927
 Qy 1122 TTATTTACTGTACACCAATGCAATTTTGAATCTGCTTAACTACTCTTCCAGGTAA 1181
 Db 1928 TTATTTACTGTACACCAATGCAATTTTGAATCTGCTTAACTACTCTTCCAGGTAA 1987
 Qy 1182 TTCCAAATTAATTTGACATCCAGTAAAGAGGCCCATCTCTTCCACTTTCCTAGTC 1241
 Db 1988 TTCCAAATTAATTTGACATCCAGTAAAGAGGCCCATCTCTTCCACTTTCCTAGTC 2047
 Qy 1242 AGTATATTCAGCAAAATTTTATGAGCCCTTACTGTGGGCAAAATCATTTGATGATAT 1301
 Db 2048 AGTATATTCAGCAAAATTTTATGAGCCCTTACTGTGGGCAAAATCATTTGATGATAT 2107
 Qy 1302 TGAGAAAAATGATTAATTCCTTATTCAGTAAATGCTACTGAGCAATCTAGTGAATC 1361
 Db 2108 TGAGAAAAATGATTAATTCCTTATTCAGTAAATGCTACTGAGCAATCTAGTGAATC 2167
 Qy 1362 ATTACAGTATGCGCTCATTTGTTGTTGAGGTGTGTTATTCATTAATATATTTTAC 1418
 Db 2168 ATTACAGTATGCGCTCATTTGTTGTTGAGGTGTGTTATTCATTAATATATTTTAC 2224
 RESULT 15
 AK054635 2473 bp mRNA linear PRI 01-AUG-2002
 LOCUS AK054635
 DEFINITION Homo sapiens cDNA FLJ30073 fis, clone ASTRO2000480.
 ACCESSION AK054635
 VERSION AK054635.1 GI:16549217
 KEYWORDS
 SOURCE
 ORGANISM
 Homo sapiens (human)
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

REFERENCE
AUTHORS
1 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
Oshima, A., Takahashi-Fujii, A., Tanase, T., Imose, N., Takeuchi, K.,
Arita, M., Mushashino, K., Yuuki, H., Hara, H., Sugiyama, T., Irie, R.,
Otsuki, T., Sato, H., Wakamatsu, A., Ishii, S., Yamamoto, J., Isono, Y.,
Kawai-Hio, Y., Saito, K., Nishikawa, T., Kimura, K., Yamashita, H.,
Masuo, K., Nakamura, Y., Sekine, M., Kikuchi, H., Kanda, K.,
Wagatsuma, M., Murakawa, K., Kanehori, K., Sugiyama, A., Kawakami, B.,
Suzuki, Y., Sugano, S., Nagahori, K., Masuo, Y., Nagai, K. and
Isozaki, T.
TITLE
JOURNAL
NEDO human cDNA sequencing project
Unpublished
2 (bases 1 to 2473)
Isozaki, T., Otsuki, T. and Sugiyama, T.
REFERENCE
AUTHORS
JOURNAL
TITLE
Direct Submission
Submitted (24-OCT-2001) Takao Isozaki, Helix Research Institute,
Genomics Laboratory; 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
(E-mail: genomics@hri.co.jp; Tel: 81-438-52-3975, Fax: 81-438-52-3986)
NEDO human cDNA sequencing project supported by Ministry of
Economy, Trade and Industry of Japan; cDNA full insert sequencing:
Research Association for Biotechnology (RAB); cDNA library
construction: Helix Research Institute (HRI) (supported by Japan
Key Technology Center etc.); 5'- & 3'-end one pass sequencing: RAB,
HRI, and Biotechnology Center, National Institute of Technology and
Evaluation; clone selection for full insert sequencing: RAB and
HRI.
FEATURES
source
1..2473
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="ASTRO200480"
/cell_type="normal astrocytes (NHA5732)"
/clone_lib="ASTRO2"
/note="Cloning vector: pME18FL3-primary culture, normal
astrocytes"

ORIGIN
Query Match 42.0%; Score 786; DB 9; Length 2473;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 836; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 562 AGACCAAACTGACGCTGGAAATACCCAGTACAGTTCCTTACTTACACTTCAAGAAAGATC 641
DB 1388 AGACCAAACTGACGCTGGAAATACCCAGTACAGTTCCTTACTTACACTTCAAGAAAGATC 1447
QY 642 TGAAGCCGGAAGAAAGAAAGAGGAGGAGTTCAGAGCAAGAGGAGTGGTGGAGGT 701
DB 1448 TGAAGCCGGAAGAAAGAAAGAGGAGGAGTTCAGAGCAAGAGGAGTGGTGGAGGT 1507
QY 702 GCTGCAGTATGAAATACGTGACGAATATTTTGACTGTGCTGTGAAAAGATTAAGATGTT 761
DB 1508 GCTGCAGTATGAAATACGTGACGAATATTTTGACTGTGCTGTGAAAAGATTAAGATGTT 1567
QY 762 ATCGAAAACCTACATGAAATTAATTGAATGCCCTTCAAGTTGAAAGTAAGCATTTTAGAGC 821
DB 1568 ATCGAAAACCTACATGAAATTAATTGAATGCCCTTCAAGTTGAAAGTAAGCATTTTAGAGC 1627
QY 822 AAATATAAGAAATTCACCTTGTACTTGTGAAACTAATCCCTAATATGAAATAGTTT 881
DB 1628 AAATATAAGAAATTCACCTTGTACTTGTGAAACTAATCCCTAATATGAAATAGTTT 1687
QY 882 AATATGATTCATGGGTAAACAGGTCCAAATTAATTAATTGAAAAGTGAATGCTGAATAT 941
DB 1688 AATATGATTCATGGGTAAACAGGTCCAAATTAATTAATTGAAAAGTGAATGCTGAATAT 1747
QY 942 CAAGGAAGACAGCATAGTCTCTTACAGGCGCTGTGGTCTGCTCAACAGTAATGG 1001
DB 1748 CAAGGAAGACAGCATAGTCTCTTACAGGCGCTGTGGTCTGCTCAACAGTAATGG 1807
QY 1002 GTGGGAAAAGGTATGCTCAATATAAAGTTCATTTTGCATTTATTTGGCAATCTTGC 1061
DB 1808 GTGGGAAAAGGTATGCTCAATATAAAGTTCATTTTGCATTTATTTGGTAAATCTTGC 1867

QY 1062 CTTTGTATTTATTTGGTGCAGTGTTCCTGCTTAATCATTTTGTGTGGCATCTGCT 1121
DB 1868 CTTTGTATTTATTTGGTGCAGTGTTCCTGCTTAATCATTTTGTGTGGCATCTGCT 1927
QY 1122 TTATTTCTTTACACACAGCAGTATTTACATCTGTCTTACTACTCTCTCCAGGTAAA 1181
DB 1928 TTATTTCTTTACACACAGCAGTATTTACATCTGTCTTACTACTCTCTCCAGGTAAA 1987
QY 1182 TTCCAAATTATTTGACATCCAGCTAAGAGGGCCCATCTCTTGCACCTCTTTCCTAGTC 1241
DB 1988 TTCCAAATTATTTGACATCCAGCTAAGAGGGCCCATCTCTTGCACCTCTTTCCTAGTC 2047
QY 1242 AGTATATTCACCAATATTTATTTAGGCCCTTACATGTGGGCAATTCATTTGTAAGTAT 1301
DB 2048 AGTATATTCACCAATATTTATTTAGGCCCTTACATGTGGGCAATTCATTTGTAAGTAT 2107
QY 1302 TGAGAAAATAGATTAATCCCTTATTCAGTAAATGTCTACTGAGCACAATCTAGTGAATC 1361
DB 2108 TGAGAAAATAGATTAATCCCTTATTCAGTAAATGTCTACTGAGCACAATCTAGTGAATC 2167
QY 1362 ATTACAGTATGCGCTCATTTGTTTGTGAGGTGTATTCATTAACAATATTTTAC 1418
DB 2168 ATTACAGTATGCGCTCATTTGTTTGTGAGGTGTATTCATTAACAATATTTTAC 2224

Search completed: July 29, 2004, 07:58:23
Job time : 7426 secs

XX 23-MAR-2000.
 PD 17-SEP-1999; 99WO-US021688.
 XX 17-SEP-1998; 98US-00156039.
 XX 22-SEP-1998; 98US-00158720.
 PR 04-NOV-1998; 98US-00186815.
 PR 08-APR-1999; 99US-0128660P.
 XX (INCY-) INCYTE PHARM INC.
 PA
 XX Tang YT, Corley NC, Guegler KJ, Gorgone GA, Patterson C,
 PI Hillman JL, Baughn MR, Lal P, Azimzai Y, Yue H, Yang J,
 XX WPI; 2000-271437/23.
 DR N-PSDB; AAA12401.
 XX
 PT New polypeptides and polynucleotides, useful for preventing and treating
 PT a disorder associated with increased or decreased expression of RNA
 PT associated proteins.
 PS
 XX Claim 1; Page 90-91; 131pp; English.
 CC The present sequence represents a human RNA-associated protein. The
 CC expression of RNA-associated proteins is closely associated with
 CC reproductive tissues, nervous tissues, cell proliferation including
 CC cancer, inflammation and immune responses, and so they may be used for
 CC diagnosis, treatment or prevention of cell proliferative,
 CC immune/inflammatory disorders, and reproductive disorders. Diseases and
 CC disorders which may be treated include actinic keratosis,
 CC atherosclerosis, arteriosclerosis, bursitis, cirrhosis, hepatitis, mixed
 CC connective tissue disease, myelofibrosis, paroxysmal nocturnal
 CC hemoglobinuria, polycythemia vera, psoriasis, primary thrombocythemia
 CC and cancers, and trauma
 XX
 SQ Sequence 216 AA;

Query Match 100.0%; Score 216; DB 3; Length 216;
 Best Local Similarity 100.0%; Pred. No. 4.6e-206;
 Matches 216; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSRYLRPPTSLFVNNVADDTSEDLREFGRYGVIVDYVPLDFTRRPRGFAYVQFND 60
 Db 1 MSRYLRPPTSLFVNNVADDTSEDLREFGRYGVIVDYVPLDFTRRPRGFAYVQFND 60
 QY 61 VDAADALHNLDRKWIICGRQIEIOFAQGDRTPNQWKAKEGNVYSSSRDYDDYDRYRGR 120
 Db 61 VDAADALHNLDRKWIICGRQIEIOFAQGDRTPNQWKAKEGNVYSSSRDYDDYDRYRGR 120
 QY 121 SRSYRRRRRSRSPYNNYRSYSPNSRPTGRPRRREALPTMIDOTAIGPSTVLLTLTQ 180
 Db 121 SRSYRRRRRSRSPYNNYRSYSPNSRPTGRPRRREALPTMIDOTAIGPSTVLLTLTQ 180
 QY 181 ERSESGKRTKEGQFKRPGKGVVLQYEVCTNLTLLV 216
 Db 181 ERSESGKRTKEGQFKRPGKGVVLQYEVCTNLTLLV 216
 RESULT 2
 ID AAY70220
 XX AAY70220 standard; protein; 216 AA.
 AC AAY70220;
 XX
 DT 06-JUN-2000 (first entry)
 XX
 DE Human RNA-associated protein-1 (RNAAP-1).
 XX
 KW RNA-associated protein; RNAAP; human; clone 399781; cytosolic;
 KW immunosuppressive; antiinflammatory; keratolytic; neuroprotective;
 KW antiarteriosclerotic; hepatotropic; antiposrotic; vincicde; anti-HIV;
 KW antiallergic; antirheumatic; antiarthritic; ophthalmological; autoimmune;

KW antimicrobial; cell proliferative disorder; inflammation; cirrhosis;
 KW actinic keratosis; bursitis; arteriosclerosis; artherosclerosis;
 KW hepatitis; myelofibrosis; primary thrombocythemia; psoriasis; cancer;
 KW mixed connective tissue disease; MCTD; HIV; uveitis; Cronh's disease;
 XX allergy; rheumatoid arthritis; parasitic infection.
 OS Homo sapiens.
 FH
 FT Modified-site
 FT 9
 FT /note= "Potential Glycosylation site"
 FT Binding-site
 FT 12..83
 FT /label= RNP-1/RM_RNA_binding_motif
 FT Modified-site
 FT 21
 FT /note= "Potential Phosphorylation site"
 FT Modified-site
 FT 47
 FT /note= "Potential Phosphorylation site"
 FT Modified-site
 FT 107
 FT /note= "Potential Phosphorylation site"
 FT Modified-site
 FT 108
 FT /note= "Potential Phosphorylation site"
 FT Modified-site
 FT 116
 FT /note= "Potential Phosphorylation site"
 FT Modified-site
 FT 129
 FT /note= "Potential Phosphorylation site"
 FT Modified-site
 FT 138
 FT /note= "Potential Phosphorylation site"
 FT Modified-site
 FT 143
 FT /note= "Potential Phosphorylation site"
 FT Modified-site
 FT 150
 FT /note= "Potential Phosphorylation site"
 FT Modified-site
 FT 161
 FT /note= "Potential Phosphorylation site"
 FT Modified-site
 FT 178
 FT /note= "Potential Phosphorylation site"
 FT Modified-site
 FT 185
 FT /note= "Potential Phosphorylation site"
 XX
 XX WO200011171-A2.
 PD 02-MAR-2000.
 XX
 PF 20-AUG-1999; 99WO-US019361.
 XX
 PR 21-AUG-1998; 98US-0097550P.
 XX 12-JAN-1999; 99US-0115633P.
 XX
 PA (INCY-) INCYTE PHARM INC.
 PI Hillman JL, Yue H, Tang YT, Corley NC, Guegler KJ, Gorgone GA;
 PI Patterson C, Baughn MR, Lal P, Bandman O, Reddy R, Azimzai Y;
 PI Shin LL, Yang J, Lu DM;
 XX WPI; 2000-237651/20.
 DR N-PSDB; AA251250.
 XX
 PT Human RNA-associated proteins useful in diagnosing, treating and
 PT preventing cell proliferative, autoimmune, inflammatory and infectious
 PT disorders.
 PS
 XX Claim 1; Page 77-78; 133pp; English.
 CC The present amino acid sequence is the human RNA-associated protein-1
 CC (RNAAP-1), identified in Incyte clone 399781, derived from PITN0702
 CC library. It is expressed in reproductive, nervous and cardiovascular
 CC tissues. It has cytosolic, immunosuppressive, keratolytic,
 CC antiinflammatory, antiarteriosclerotic, hepatotropic, vincicde,
 CC neuroprotective, antiposrotic, anti-HIV, antiallergic, antirheumatic,
 CC antiahrthic, ophthalmological and antimicrobial activity. RNAAP
 CC antibodies are useful for diagnosis of diseases associated with altered
 CC expression or activity of RNAAP. It is used to treat cell proliferative,
 CC autoimmune, inflammatory and infectious disorders, like actinic
 CC keratosis, bursitis, arteriosclerosis, atherosclerosis, cirrhosis,

CC hepatitis, myelofibrosis, mixed connective tissue disease (MCTD),
 CC psoriasis, primary thrombocytopenia and cancer, HIV, allergies, rheumatoid
 CC arthritis, uveitis, Crohn's disease, and bacterial, viral and parasitic
 CC infections

SQ Sequence 216 AA;

Query Match 100.0%; Score 216; DB 3; Length 216;
 Best Local Similarity 100.0%; Pred. No. 4.6e-206;
 Matches 216; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSRYLRPNTSLFVNVADDTRESEDLRRFGYGPVIVVYPLDFTYRRRGFAVQFED 60
 DB 1 MSRYLRPNTSLFVNVADDTRESEDLRRFGYGPVIVVYPLDFTYRRRGFAVQFED 60
 QY 61 VRDAEDALHNLDRKWIICGRQIEICFAQGDRTKPNOMKAKGRNVSSSRDYDDYRRSR 120
 DB 61 VRDAEDALHNLDRKWIICGRQIEICFAQGDRTKPNOMKAKGRNVSSSRDYDDYRRSR 120
 QY 121 SRSYERRRSRSPFDYNNRSGSPNSRPTGPRRRRAIPMIDQTAAGIPSTVLLTTLQ 180
 DB 121 SRSYERRRSRSPFDYNNRSGSPNSRPTGPRRRRAIPMIDQTAAGIPSTVLLTTLQ 180
 QY 181 ERSSEGRKTKEGQFRPKGKMKVLQYECNTLITLV 216
 DB 181 ERSSEGRKTKEGQFRPKGKMKVLQYECNTLITLV 216

RESULT 3
 AAY70228
 ID AAY70228 standard; protein, 183 AA.

XX AAY70228;
 XX 06-JUN-2000 (first entry)

DE Human RNA-associated protein-9 (RNAAP-9).

XX RNA-associated protein; RNAAP; human; clone 934406; cytosolic;
 XX immunosuppressive; anti-inflammatory; antiproliferative;
 XX antiarteriosclerotic; hepatotropic; antiproliferative; anti-HIV;
 XX antiallergic; antineoplastic; antirheumatic; antiparasitic;
 XX antimicrobial; cell proliferative disorder; inflammation; cirrhosis;
 XX actinic keratosis; bursitis; arteriosclerosis; artherosclerosis;
 XX hepatitis; myelofibrosis; primary thrombocytopenia; psoriasis; cancer;
 XX mixed connective tissue disease; MCTD; HIV; uveitis; Crohn's disease;
 XX allergy; rheumatoid arthritis; parasitic infection.

OS Homo sapiens.

XX Homo sapiens.

XX Key Location/Qualifiers
 XX Modified-site 9 /note= "Potential glycosylation site"

XX Domain 12..83 /label= RNA_recognition_motif

XX Modified-site 21 /note= "Potential phosphorylation site"

XX Region 30..81 /note= "Signature sequence of RNP-1"

XX Modified-site 47 /note= "Potential phosphorylation site"

XX Modified-site 107 /note= "Potential phosphorylation site"

XX Modified-site 108 /note= "Potential phosphorylation site"

XX Modified-site 116 /note= "Potential phosphorylation site"

XX Modified-site 129 /note= "Potential phosphorylation site"

XX Modified-site 138 /note= "Potential phosphorylation site"

XX Modified-site 143 /note= "Potential phosphorylation site"

XX Modified-site /note= "Potential phosphorylation site"

FT Modified-site 150 /note= "Potential phosphorylation site"
 FT Modified-site 158 /note= "Potential phosphorylation site"
 FT Modified-site 160 /note= "Potential phosphorylation site"
 FT Modified-site 166 /note= "Potential phosphorylation site"
 FT Modified-site 179 /note= "Potential glycosylation site"
 FT Modified-site 180 /note= "Potential phosphorylation site"
 FT Modified-site /note= "Potential phosphorylation site"

PN WO200011171-A2.

PD 02-MAR-2000.

PF 20-AUG-1999; 99WO-US019361.

PR 21-AUG-1998; 98US-0097550P.

PR 12-JAN-1999; 99US-0115639P.

PA (INCY-) INCYTE PHARM INC.

PI Hillman JL, Yue H, Tang YT, Corley NC, Guegler KJ, Gorgone GA;
 PI Patterson C, Baughn MR, Lal P, Bandman O, Reddy R, Azimzai Y;
 PI Shih JL, Yang J, Lu DM;

DR WPI, 2000-237651/20.

DR N-PSDB; AAZ51258.

PT Human RNA-associated proteins useful in diagnosing, treating and
 PT preventing cell proliferative, autoimmune, inflammatory and infectious
 PT disorders.

PS Claim 1; Page 86; 123p; English.

XX The present amino acid sequence is the human RNA-associated protein-9
 CC (RNAAP-9), identified in Incyte clone 934406, derived from CERNO101
 CC library. It is expressed in reproductive, nervous, cardiovascular and
 CC haematopoietic/immune tissues. It has cytosolic, immunosuppressive,
 CC antineoplastic, antiarteriosclerotic, hepatotropic, keratolytic,
 CC neuroprotective, antiparasitic, anti-HIV, antiallergic, antirheumatic,
 CC vitreous, antirheumatic, ophthalmological and antimicrobial activity.
 CC RNAAP antibodies are useful for diagnosis of diseases associated with
 CC altered expression or activity of RNAAP. It is used to treat cell
 CC proliferative, autoimmune, inflammatory and infectious disorders, like
 CC actinic keratosis, bursitis, arteriosclerosis, artherosclerosis,
 CC cirrhosis, hepatitis, myelofibrosis, mixed connective tissue disease
 CC (MCTD), psoriasis, primary thrombocytopenia and cancer, HIV, allergies,
 CC rheumatoid arthritis, uveitis, Crohn's disease, and bacterial, viral and
 CC parasitic infections

XX Sequence 183 AA;

QY Query Match 71.8%; Score 155; DB 3; Length 183;
 Best Local Similarity 100.0%; Pred. No. 1.5e-145;

Matches 155; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSRYLRPNTSLFVNVADDTRESEDLRRFGYGPVIVVYPLDFTYRRRGFAVQFED 60
 DB 1 MSRYLRPNTSLFVNVADDTRESEDLRRFGYGPVIVVYPLDFTYRRRGFAVQFED 60

QY 61 VRDAEDALHNLDRKWIICGRQIEICFAQGDRTKPNOMKAKGRNVSSSRDYDDYRRSR 120
 DB 61 VRDAEDALHNLDRKWIICGRQIEICFAQGDRTKPNOMKAKGRNVSSSRDYDDYRRSR 120

QY 121 SRSYERRRSRSPFDYNNRSGSPNSRPTGPRRRRAIPMIDQTAAGIPSTVLLTTLQ 155
 DB 121 SRSYERRRSRSPFDYNNRSGSPNSRPTGPRRRRAIPMIDQTAAGIPSTVLLTTLQ 155

RESULT 4

AA92806
ID AAB92806 standard; protein, 183 AA.
XX
AC AAB92806;
XX
DT 26-JUN-2001 (first entry)
XX
DE Human protein sequence SEQ ID NO:11320.
DE
KW Human; primer; detection; diagnosis; antisense therapy; gene therapy.
XX
OS Homo sapiens.
XX
EN EPI074617-A2.
PD
PD 07-FEB-2001.
PF 28-JUL-2000; 2000EP-00116126.
XX
XX 29-JUL-1999; 99JP-00248036.
PR 27-AUG-1999; 99JP-00300253.
PR 11-JAN-2000; 2000JP-00118776.
PR 02-MAY-2000; 2000JP-00183767.
PR 09-JUN-2000; 2000JP-00241899.
XX
XX (HELI-) HELIX RES INST.
PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX
XX WPI; 2001-318749/34.
XX
XX Primer sets for synthesizing polynucleotides, particularly the 5602 full-
PT length cDNAs defined in the specification, and for the detection and/or
PT diagnosis of the abnormality of the proteins encoded by the full-length
PT cDNAs.
XX
XX Claim 8; SEQ ID NO 11320; 2537bp + Sequence listing; English.
XX
XX The present invention describes primer sets for synthesizing 5602 full-
CC length cDNAs defined in the specification. Where a primer set comprises:
CC (a) an oligo-dT primer and an oligonucleotide complementary to the
CC complementary strand of a polynucleotide which comprises one of the 5602
CC nucleotide sequences defined in the specification, where the
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC of an oligonucleotide comprising a sequence complementary to the
CC complementary strand of a polynucleotide which comprises a 5'-end
CC sequence and an oligonucleotide comprising a sequence complementary to a
CC polynucleotide which comprises a 3'-end sequence, where the
CC oligonucleotide comprises at least 15 nucleotides and the combination of
CC the 5'-end sequence/3'-end sequence is selected from those defined in the
CC specification. The primer sets can be used in antisense therapy and in
CC gene therapy. The primers are useful for synthesizing polynucleotides,
CC particularly full-length cDNAs. The primers are also useful for the
CC detection and/or diagnosis of the abnormality of the proteins encoded by
CC the full-length cDNAs. The primers allow obtaining of the full-length
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to AAB95893
CC represent human amino acid sequences; and AAH13629 to AAH13632 represent
CC oligonucleotides, all of which are used in the exemplification of the
CC present invention
XX
XX Sequence 183 AA;
SQ
Query Match 71.8%; Score 155; DB 4; Length 183;
Best Local Similarity 100.0%; Pred. No. 1.5e-145;
Matches 155; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MSRYLRPNTSIFRNVAADTRESDDRFRGGRGPIVDYVYVPLDFTTRPRGAYVQFED 60
DB 1 MSRYLRPNTSIFRNVAADTRESDDRFRGGRGPIVDYVYVPLDFTTRPRGAYVQFED 60
QY 61 VDAEDALNLDKRWICGRQIEIQFAGDRTKPNQAKKGRNVYSSSRDYDRYRSR 120

DB 61 VDAEDALNLDKRWICGRQIEIQFAGDRTKPNQAKKGRNVYSSSRDYDRYRSR 120
QY 121 SRSYERRRSRSPDYNYSRSPNSRPTGRPR 155
DB 121 SRSYERRRSRSPDYNYSRSPNSRPTGRPR 155
RESULT 5
AAB93180
ID AAB93180 standard; protein, 261 AA.
XX
AC AAB93180;
XX
DT 26-JUN-2001 (first entry)
XX
DE Human protein sequence SEQ ID NO:12124.
DE
KW Human; primer; detection; diagnosis; antisense therapy; gene therapy.
XX
OS Homo sapiens.
XX
EN EPI074617-A2.
PD
PD 07-FEB-2001.
PF 28-JUL-2000; 2000EP-00116126.
XX
XX 29-JUL-1999; 99JP-00248036.
PR 27-AUG-1999; 99JP-00300253.
PR 11-JAN-2000; 2000JP-00118776.
PR 02-MAY-2000; 2000JP-00183767.
PR 09-JUN-2000; 2000JP-00241899.
XX
XX (HELI-) HELIX RES INST.
PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX
XX WPI; 2001-318749/34.
XX
XX Primer sets for synthesizing polynucleotides, particularly the 5602 full-
PT length cDNAs defined in the specification, and for the detection and/or
PT diagnosis of the abnormality of the proteins encoded by the full-length
PT cDNAs.
XX
XX Claim 8; SEQ ID NO 12124; 2537bp + Sequence listing; English.
XX
XX The present invention describes primer sets for synthesizing 5602 full-
CC length cDNAs defined in the specification. Where a primer set comprises:
CC (a) an oligo-dT primer and an oligonucleotide complementary to the
CC complementary strand of a polynucleotide which comprises one of the 5602
CC nucleotide sequences defined in the specification, where the
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC of an oligonucleotide comprising a sequence complementary to the
CC complementary strand of a polynucleotide which comprises a 5'-end
CC sequence and an oligonucleotide comprising a sequence complementary to a
CC polynucleotide which comprises a 3'-end sequence, where the
CC oligonucleotide comprises at least 15 nucleotides and the combination of
CC the 5'-end sequence/3'-end sequence is selected from those defined in the
CC specification. The primer sets can be used in antisense therapy and in
CC gene therapy. The primers are useful for synthesizing polynucleotides,
CC particularly full-length cDNAs. The primers are also useful for the
CC detection and/or diagnosis of the abnormality of the proteins encoded by
CC the full-length cDNAs. The primers allow obtaining of the full-length
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to AAB95893
CC represent human amino acid sequences; and AAH13629 to AAH13632 represent
CC oligonucleotides, all of which are used in the exemplification of the
CC present invention
XX
XX Sequence 261 AA;
SQ

Query Match 67.6%; Score 146; DB 4; Length 261;
Best Local Similarity 100.0%; Pred. No. 1.8e-136;
Matches 146; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSRYLRPPTSLFVNNAVDDTSEDLRRFGRGPIVDYVPLDFTYRPRGFAVQFED 60
DB 1 MSRYLRPPTSLFVNNAVDDTSEDLRRFGRGPIVDYVPLDFTYRPRGFAVQFED 60
QY 61 VRDAEDALHNLDKWKICGRQIEIOPAGDRKTPNOMKAKEGNVSSRYDDYDRYRRSR 120
DB 61 VRDAEDALHNLDKWKICGRQIEIOPAGDRKTPNOMKAKEGNVSSRYDDYDRYRRSR 120
QY 121 SRSYERRRSRSPFNRRSYSPRN 146
DB 121 SRSYERRRSRSPFNRRSYSPRN 146

RESULT 6
ABG00682
ID ABG00682 standard; protein, 198 AA.
AC ABG00682;
XX 13-FEB-2002 (first entry)
DT Novel human diagnostic protein #673.
DE
XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder.
OS Homo sapiens.
XX W0200175067-A2.
XX 11-OCT-2001.
PD 30-MAR-2001; 2001WO-US008631.
PF 31-MAR-2000; 2000US-00540217.
PR 23-AUG-2000; 2000US-00649167.
XX (HYSE-) HYSEQ INC.
PA Dmanac RT, Liu C, Tang YT;
PI WPI; 2001-639362/73.
DR N-PSDB; AAS64869.
XX

New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity.
XX
XX Claim 20; SEQ ID NO 31041; 103pp; English.

The invention relates to isolated polynucleotide (I) and polypeptide (II)
CC sequences. (I) is useful as hybridisation probes, polymerase chain
CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
CC and in recombinant production of (II). The polynucleotides are also used
CC in diagnostics as expressed sequence tags for identifying expressed
CC genes. (I) is useful in gene therapy techniques to restore normal
CC activity of (II) or to treat disease states involving (II). (II) is
CC useful for generating antibodies against it, detecting or quantitating a
CC polypeptide in tissue, as molecular weight markers and as a food
CC supplement. (II) and its binding partners are useful in medical imaging
CC of sites expressing (II). (I) and (II) are useful for treating disorders
CC involving aberrant protein expression or biological activity. The
CC polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG00010-ABG30377 represent novel human diagnostic
CC amino acid sequences of the invention. Note: The sequence data for this

CC patent did not appear in the printed specification, but was obtained in
CC electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX

SQ Sequence 198 AA;

Query Match 66.7%; Score 144; DB 4; Length 198;
Best Local Similarity 100.0%; Pred. No. 1.4e-134;
Matches 144; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 12 LFNNAVDDTSEDLRRFGRGPIVDYVPLDFTYRPRGFAVQFEDVRDAEDALHNL 71
DB 12 LFNNAVDDTSEDLRRFGRGPIVDYVPLDFTYRPRGFAVQFEDVRDAEDALHNL 71
QY 72 DRKWCGRQIEIOPAGDRKTPNOMKAKEGNVSSRYDDYDRYRRSRYSYERRRSR 131
DB 72 DRKWCGRQIEIOPAGDRKTPNOMKAKEGNVSSRYDDYDRYRRSRYSYERRRSR 131
QY 132 RSPDYNYRRSYSPNRRPTGRPRR 155
DB 132 RSPDYNYRRSYSPNRRPTGRPRR 155

RESULT 7
ABG15228
ID ABG15228 standard; protein, 238 AA.
XX
XX ABG15228;
AC
XX 18-FEB-2002 (first entry)
DT Novel human diagnostic protein #15219.
DE
XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder.
OS Homo sapiens.
XX W0200175067-A2.
XX 11-OCT-2001.
PD 30-MAR-2001; 2001WO-US008631.
PF 31-MAR-2000; 2000US-00540217.
PR 23-AUG-2000; 2000US-00649167.
XX (HYSE-) HYSEQ INC.
PA Dmanac RT, Liu C, Tang YT;
PI WPI; 2001-639362/73.
DR N-PSDB; AAS79415.
XX

New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity.
XX
XX Claim 20; SEQ ID NO 45587; 103pp; English.

The invention relates to isolated polynucleotide (I) and polypeptide (II)
CC sequences. (I) is useful as hybridisation probes, polymerase chain
CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
CC and in recombinant production of (II). The polynucleotides are also used
CC in diagnostics as expressed sequence tags for identifying expressed
CC genes. (I) is useful in gene therapy techniques to restore normal
CC activity of (II) or to treat disease states involving (II). (II) is
CC useful for generating antibodies against it, detecting or quantitating a
CC polypeptide in tissue, as molecular weight markers and as a food
CC supplement. (II) and its binding partners are useful in medical imaging
CC of sites expressing (II). (I) and (II) are useful for treating disorders
CC involving aberrant protein expression or biological activity. The

CC polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG00010-ABG30377 represent novel human diagnostic
CC amino acid sequences of the invention. Note: The sequence data for this
CC patent did not appear in the printed specification, but was obtained in
CC electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 238 AA;
Query Match 66.7%; Score 144; DB 4; Length 238;
Best Local Similarity 100.0%; Pred. No. 1.6e-134; Indels 0; Gaps 0;
Matches 144; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 12 LFVNNVADDTRESEDLRRFGRGYPIDVYVPLDFYTRRPRGFAYVQPEDVADALHNL 71
DB 49 LFVNNVADDTRESEDLRRFGRGYPIDVYVPLDFYTRRPRGFAYVQPEDVADALHNL 108
QY 72 DRKMICGQIIEIOFAGQDRKTPNQMKAKGKNVYSSSYDYDRIYRRSRSSYERRRSRS 131
DB 109 DRKMICGQIIEIOFAGQDRKTPNQMKAKGKNVYSSSYDYDRIYRRSRSSYERRRSRS 168
QY 132 RSPDYNNRYSYSPNSRPTGRPRR 155
DB 169 RSPDYNNRYSYSPNSRPTGRPRR 192
RESULT 8
ABG00681
ID ABG00681 standard; protein; 118 AA.
AC ABG00681;
XX
DT 13-FEB-2002 (first entry)
XX
DE Novel human diagnostic protein #672.
XX
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder.
XX
OS Homo sapiens.
XX
PN WO200175067-A2.
XX
PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001WO-US008631.
XX
PR 31-MAR-2000; 2000US-00540217.
XX
PR 23-AUG-2000; 2000US-00649167.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Dmanac RT, Liu C, Tang YT;
XX
PI WPI; 2001-639362/73.
XX
DR N-PSDB; AAS64868.
XX
PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity.
XX
PS Claim 20; SEQ ID NO 31040; 103bp; English.
XX
XX The invention relates to isolated polynucleotide (I) and polypeptide (II)
XX sequences. (I) is useful as hybridisation probes, polymerase chain
XX reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
XX and in recombinant production of (II). The polynucleotides are also used
XX in diagnostics as expressed sequence tags for identifying expressed
XX genes. (I) is useful in gene therapy techniques to restore normal

CC activity of (II) or to treat disease states involving (II). (II) is
CC useful for generating antibodies against it, detecting or quantitating a
CC polypeptide in tissue, as molecular weight markers and as a food
CC supplement. (II) and its binding partners are useful in medical imaging
CC of sites expressing (II). (I) and (II) are useful for treating disorders
CC involving aberrant protein expression or biological activity. The
CC polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG00010-ABG30377 represent novel human diagnostic
CC amino acid sequences of the invention. Note: The sequence data for this
CC patent did not appear in the printed specification, but was obtained in
CC electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 118 AA;
Query Match 44.4%; Score 96; DB 4; Length 118;
Best Local Similarity 100.0%; Pred. No. 4.3e-87; Indels 0; Gaps 0;
Matches 96; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MSRTLRPPNTSLFVRNNAADTRESDELRRFGRGYPIDVYVPLDFYTRRPRGFAYVQPED 60
DB 1 MSRTLRPPNTSLFVRNNAADTRESDELRRFGRGYPIDVYVPLDFYTRRPRGFAYVQPED 60
QY 61 VRDAEDALHNLDRKMICGQIIEIOFAGQDRKTPNQM 96
DB 61 VRDAEDALHNLDRKMICGQIIEIOFAGQDRKTPNQM 96
RESULT 9
ABG15227
ID ABG15227 standard; protein; 130 AA.
AC ABG15227;
XX
DT 18-FEB-2002 (first entry)
XX
DE Novel human diagnostic protein #15218.
XX
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder.
XX
OS Homo sapiens.
XX
PN WO200175067-A2.
XX
PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001WO-US008631.
XX
PR 31-MAR-2000; 2000US-00540217.
XX
PR 23-AUG-2000; 2000US-00649167.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Dmanac RT, Liu C, Tang YT;
XX
PI WPI; 2001-639362/73.
XX
DR N-PSDB; AAS79414.
XX
PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity.
XX
PS Claim 20; SEQ ID NO 45586; 103bp; English.
XX
XX The invention relates to isolated polynucleotide (I) and polypeptide (II)
XX sequences. (I) is useful as hybridisation probes, polymerase chain
XX reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
XX and in recombinant production of (II). The polynucleotides are also used

in diagnostics as expressed sequence tags for identifying expressed genes. (I) is useful in gene therapy techniques to restore normal activity of (II) or to treat disease states involving (II). (II) is useful for generating antibodies against it, detecting or quantitating a polypeptide in tissue, as molecular weight markers and as a food supplement. (II) and its binding partners are useful in medical imaging of sites expressing (II). (I) and (II) are useful for treating disorders involving aberrant protein expression or biological activity. The polypeptide and polynucleotide sequences have applications in diagnostics, forensics, gene mapping, identification of mutations responsible for genetic disorders or other traits to assess biodiversity and to produce other types of data and products dependent on DNA and amino acid sequences. ABG00010-ABG3037 represent novel human diagnostic parent did not appear in the invention. Note: The sequence data for this electronic format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences

Sequence 130 AA;

Query Match 44.4%; Score 96; DB 4; Length 130;
Best Local Similarity 100.0%; Pred. No. 4.6e-87;
Matches 96; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSRYLRPNTSLFVRNVADTRSEDLRRERGRYGVIVVYPLDPTTRPRGFAYVQPED 60
DB 7 MSRYLRPNTSLFVRNVADTRSEDLRRERGRYGVIVVYPLDPTTRPRGFAYVQPED 66
QY 61 VRDAEDALHNDLRKWCGRQIEIQFAGCDRKTNPQM 96
DB 67 VRDAEDALHNDLRKWCGRQIEIQFAGCDRKTNPQM 102

RESULT 10

AA02102 standard; protein; 92 AA.

AA02102;

06-OCT-2000 (first entry)

Human secreted protein, SEQ ID NO: 6183.

Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation; gene therapy; chromosome mapping.

Homo sapiens.

EP1033401-A2.

06-SEP-2000.

21-FEB-2000; 2000EP-00200610.

26-FEB-1999; 99US-0122487P.

(GEST) GENSET.

Dumas Milne Edwards J, Duclert A, Giordano J;

WPI; 2000-500381/45.

N-PDB; AAC02108.

New nucleic acid that is a 5' expressed sequence tag (5' EST) for obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for diagnostic, forensic, gene therapy and chromosome mapping procedures.

Claim 13; SEQ ID NO 6183; 71bp + Sequence Listing; English.

The present sequence is a polypeptide encoded by one of a large number of 5' ESTs derived from mRNAs encoding secreted proteins. The 5' ESTs were prepared from total human RNAs or polyA+ RNAs derived from 30 different tissues. EST sequences usually correspond mainly to the 3' untranslated

region (UTR) of the mRNA because they are often obtained from oligo-dT primed cDNA libraries. Such ESTs are not well suited for isolating cDNA sequences derived from the 5' ends of mRNAs and even in those cases where longer cDNA sequences have been obtained, the full 5' UTR is rarely included. 5' ESTs are derived from mRNAs with intact 5' ends and can therefore be used to obtain full length cDNAs and genomic DNAs. 5' ESTs are also used in diagnostic, forensic, gene therapy and chromosome mapping procedures. They are used to obtain upstream regulatory sequences and to design expression and secretion vectors

Sequence 92 AA;

Query Match 42.6%; Score 92; DB 3; Length 92;
Best Local Similarity 100.0%; Pred. No. 3.2e-83;
Matches 92; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSRYLRPNTSLFVRNVADTRSEDLRRERGRYGVIVVYPLDPTTRPRGFAYVQPED 60
DB 1 MSRYLRPNTSLFVRNVADTRSEDLRRERGRYGVIVVYPLDPTTRPRGFAYVQPED 60
QY 61 VRDAEDALHNDLRKWCGRQIEIQFAGCDRKT 92
DB 61 VRDAEDALHNDLRKWCGRQIEIQFAGCDRKT 92

RESULT 11

AAE36184 standard; protein; 261 AA.

AAE36184;

26-JUN-2003 (first entry)

Human CGPD-15 protein.

Human; cell growth; differentiation; cell proliferative disorder; CGPD; death; cancer; cirrhosis; hepatitis; psoriasis; developmental disorder; cataract; anaemia; Cushing's syndrome; neurological disorder; epilepsy; Huntington's disease; Parkinson's disease; Crohn's disease; transgenic; Goodpasture's syndrome; inflammatory disorder; reproductive disorder; rheumatoid arthritis; Grave's disease; dementia; autoimmune disorder; infertility; endometriosis; placenta disorder; preeclampsia; diabetes; metabolic disorder; tumour; cancer; choriocarcinoma; chronic villitis; infertility; obesity; transgenic animal; gene therapy.

Homo sapiens.

Key Location/Qualifiers

Domain 12..83 /note="RNA recognition motif"

WO2002102310-A2.

27-DEC-2002.

12-JUN-2002; 2002WO-US018834.

15-JUN-2001; 2001US-0298617P.
21-JUN-2001; 2001US-0300376P.
29-JUN-2001; 2001US-0301873P.
09-JUL-2001; 2001US-0304053P.
13-JUL-2001; 2001US-0305330P.
13-JUL-2001; 2001US-0305361P.
13-JUL-2001; 2001US-0305370P.

(INCY-) INCYTE GENOMICS INC.

Yue H, Lu DM, Hafalia Afa, Arvizu CS, Ramkumar J, Tang YT;
Khan FA, Greene BD, Richardson TW, Yang J, Ison CH, Warren BA;
Elliot VS, Emerling BM, Gorvay AE, Lee EA, Griffin JA;
Zebajadjan Y, Svarnakar A, Lal PG, Baughn MR, Tran UK, Lee S;
Foreythe IV, Au-Young JK, Coleman IM;


```

XX 28-MAR-2001; 2001US-0279619P.
PR 29-MAR-2001; 2001US-0280067P.
PR 29-MAR-2001; 2001US-0280068P.
PR 16-MAY-2001; 2001US-0291280P.
PR 17-MAY-2001; 2001US-0291829P.
PR 19-JUN-2001; 2001US-0291849P.
PR 20-JUN-2001; 2001US-0299428P.
PR 20-JUN-2001; 2001US-0299776P.
PR 20-JUN-2001; 2001US-0300001P.
XX
XX (INCYTE GENOMICS INC.
XX
XX Daffo A, Jones AL, Tran AB, Dahl CR, Gietzen D, Chinn J;
XX Dufour GE, Hillman JL, Yu JY, Tuason O, Yap PR, Amshey SR;
XX Daugherty SC, Dam TC, Liu TF, Nguyen DA, Kleeefeld Y, Gerstin EH;
XX Peralta CH, David MH, Lewis SA, Chen AJ, Panzer SR, Harris B;
XX Flores V, Marwaha R, Lo A, Lan RY, Urashika ME;
XX
XX WPI; 2003-058431/05.
XX N-PSDB; ABX34763.
XX
XX New purified disease detection and treatment molecule proteins and
XX polynucleotides, useful for diagnosing, treating or preventing cancers
XX (e.g. leukemia or sarcoma), anemia, Crohn's disease, AIDS, osteoporosis
XX or hepatitis.
XX
XX Claim 27; SEQ ID NO 720; 339pp + Sequence Listing; English.
XX
XX This invention describes a novel disease detection and treatment molecule
XX polypeptide (MDPT) which has anti-inflammatory, immunosuppressive,
XX osteopathic, cyostatic, anti-HIV, haemostatic, nephrotropic,
XX antiaiemic, antipsoritic and hepatotropic activity. The polynucleotides
XX and the polypeptides of the invention can be used for gene therapy,
XX protein replacement therapy and are useful for treating a variety of
XX diseases or conditions. These polypeptides or polynucleotides are
XX particularly useful for diagnosing, treating or preventing cell
XX proliferative disorders (e.g. cancers including adenocarcinoma,
XX leukemia, lymphoma, melanoma, myeloma or sarcoma), anaemia, Crohn's
XX disease, acquired immunodeficiency syndrome (AIDS), Goodpasture's
XX syndromes, inflammation, osteoporosis, thrombocytopenia, psoriasis or
XX hepatitis; ABU11450-ABU11845 represent the MDPT polynucleotides encoded
XX by ABU11450-ABU11845, described in the disclosure of the invention. NOTE:
XX The sequence data for this patent did not form part of the printed
XX specification, but was obtained in electronic format from WIPO at
XX ftp.wipo.int/pub/published_pct_sequences
XX
XX Sequence 280 AA:
XX
XX Query Match 12.5%; Score 27; DB 6; Length 280;
XX Best Local Similarity 100.0%; Pred. No. 3e-18;
XX Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX 42 PLDFYTRRRGFAVVOEFEDVDAEDAL 68
XX |||||
XX 61 PLDFYTRRRGFAVVOEFEDVDAEDAL 87
XX
XX RESULT 14
XX ID AAM39480 standard; protein; 287 AA.
XX
XX AAM39480;
XX
XX 22-OCT-2001 (first entry)
XX
XX Human polypeptide SEQ ID NO 2625.
XX
XX Human; nootropic; immunosuppressant; cyostatic; gene therapy; cancer;
XX peripheral nervous system; neuropathy; central nervous system; CNS;
XX Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
XX amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
XX chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
XX

```

```

XX leukemia.
XX
XX Homo sapiens.
XX
XX WO200153312-A1.
XX
XX 26-JUL-2001.
XX
XX 26-DEC-2000; 2000WO-US034263.
XX
XX 23-DEC-1999; 99US-00471275.
XX 21-JAN-2000; 2000US-0048725.
XX 25-APR-2000; 2000US-00552317.
XX 20-JUN-2000; 2000US-00598042.
XX 19-JUL-2000; 2000US-00620312.
XX 03-AUG-2000; 2000US-00653450.
XX 14-SEP-2000; 2000US-00662191.
XX 19-OCT-2000; 2000US-00693036.
XX 29-NOV-2000; 2000US-00727344.
XX
XX (HYSE-) HYSEQ INC.
XX
XX Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
XX Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J, Zhao QA;
XX Zhou F, Goodrich R, Drmanac RT;
XX
XX WPI; 2001-442253/47.
XX N-PSDB; AA158636.
XX
XX Novel nucleic acids and polypeptides, useful for treating disorders such
XX as central nervous system injuries.
XX
XX Example 4; SEQ ID NO 2625; 10078pp; English.
XX
XX The invention relates to human nucleic acids (AA157798-AA161369) and the
XX encoded polypeptides (AAM38642-AAM42213) with nootropic,
XX immunosuppressant and cyostatic activity. The polynucleotides are useful
XX in gene therapy. A composition containing a polypeptide or polynucleotide
XX of the invention may be used to treat diseases of the peripheral nervous
XX system, such as peripheral nervous injuries, peripheral neuropathy and
XX Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
XX lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
XX utilisation of the activities such as: Immune system suppression,
XX Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
XX and thrombolytic activity, cancer diagnosis and therapy, drug screening,
XX assays for receptor activity, arthritis and inflammation, leukemias and
XX C.N.S disorders. Note: The sequence data for this patent did not form
XX part of the printed specification
XX
XX Sequence 287 AA:
XX
XX Query Match 12.5%; Score 27; DB 4; Length 287;
XX Best Local Similarity 100.0%; Pred. No. 3.1e-18;
XX Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX 42 PLDFYTRRRGFAVVOEFEDVDAEDAL 68
XX |||||
XX 68 PLDFYTRRRGFAVVOEFEDVDAEDAL 94
XX
XX RESULT 15
XX ID ABB61853 standard; protein; 706 AA.
XX
XX ABB61853;
XX
XX 26-MAR-2002 (first entry)
XX
XX Drosophila melanogaster polypeptide SEQ ID NO 12351.
XX
XX Drosophila; developmental biology; cell signalling; insecticide;
XX pharmaceutical.
XX

```

OS Drosophila melanogaster.
 XX
 XX WO200171042-A2.
 XX
 XX 27-SEP-2001.
 PD
 XX 23-MAR-2001; 2001WO-US009231.
 PF
 XX 23-MAR-2000; 2000US-0191637P.
 PR 11-JUL-2000; 2000US-00614150.
 PR
 XX (PEKE) PE CORP NY.
 PA
 XX Venter JC, Adams M, Li PWD, Myers EW;
 PI
 XX WPI; 2001-656860/75.
 DR
 DR N-PSDB; ABL05956.
 XX
 XX New isolated nucleic acid detection reagent for detecting 1000 or more
 PT genes from Drosophila and for elucidating cell signaling and cell-cell
 PT interactions.
 PT
 PS Disclosure; SEQ ID NO 12351; 21pp + Sequence Listing; English.
 XX
 CC The invention relates to an isolated nucleic acid detection reagent
 CC capable of detecting 1000 or more genes from Drosophila. The invention is
 CC useful in developmental biology and in elucidating cell signaling and
 CC cell-cell interactions in higher eukaryotes for the development of
 CC insecticides, therapeutics and pharmaceutical drugs. The invention
 CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
 CC sequences (ABL01640-ABL16175) and the encoded proteins (ABBS7737-
 CC ABB72072). The sequence data for this patent did not form part of the
 CC printed specification, but was obtained in electronic format directly
 CC from WIPO at ftp.wipo.int/pub/published_pct_sequences
 XX
 SQ Sequence 706 AA;

 QY 117 RRSRSRSYERRR 128
 |||||||||
 Db 145 RRSRSRSYERRR 156

Search completed: July 22, 2004, 17:18:50
 Job time : 56 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 22, 2004, 17:17:52 ; Search time 16 Seconds
(without alignments)
1298.586 Million cell updates/sec

Title: US-09-787-491B-1
Perfect score: 216
Sequence: 1 MSRLRPNTSLFVRNVD.....FKGKRYLYEYCNILTLV 216
Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 283366 seqs, 96191526 residues
Word size : 0

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database :

PIR 78:*
1: p1r1:*
2: p1r2:*
3: p1r3:*
4: p1r4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	9	4.2	299	T12483	hypothetical prote
2	9	4.2	347	T06671	hypothetical prote
3	9	4.2	524	I55595	splicing factor -
4	8	3.7	55	VHNVAC	nucleocapsid prote
5	8	3.7	58	S34045	protamine - North
6	8	3.7	65	VHNVBM	nucleocapsid prote
7	8	3.7	136	D35846	transformer-2 sex-
8	8	3.7	163	I38191	nucleic acid bindi
9	8	3.7	179	T34145	hypothetical prote
10	8	3.7	196	E84638	probable R52p22 sp
11	8	3.7	199	S26047	transformer sex-de
12	8	3.7	201	S26045	sex-determining pr
13	8	3.7	221	S26045	PR264 protein - ch
14	8	3.7	221	B42701	splicing factor, a
15	8	3.7	238	A57198	major polyhedrai c
16	8	3.7	252	D72866	transformer-2 sex-
17	8	3.7	256	A32373	hypothetical prote
18	8	3.7	264	A31638	RNA-binding protei
19	8	3.7	281	T26084	hypothetical prote
20	8	3.7	286	T09704	hypothetical prote
21	8	3.7	288	S68798	hypothetical prote
22	8	3.7	299	T25596	hypothetical prote
23	8	3.7	312	T26085	hypothetical prote
24	8	3.7	315	T41868	hypothetical prote
25	8	3.7	344	S59043	splicing factor SR
26	8	3.7	345	T33906	hypothetical prote
27	8	3.7	363	B84832	hypothetical prote
28	8	3.7	378	S04336	UI snRNP 70K prote
29	8	3.7	448	A36311	70K UI small nucle

30	8	3.7	471	2	S02016	UI snRNP 70K prote
31	8	3.7	488	2	S36570	E2 protein - human
32	8	3.7	568	2	B96634	probable U2 snRNP
33	8	3.7	614	2	A25707	UI snRNP 70K prote
34	8	3.7	672	2	T21469	hypothetical prote
35	8	3.7	708	2	T25725	hypothetical prote
36	8	3.7	886	2	A32755	beta-amyloid-like
37	8	3.7	891	2	G84693	probable proline-r
38	8	3.7	919	2	S28179	alpha-amyliase (EC
39	8	3.7	1032	2	A57514	RNA helicase HEH11
40	8	3.7	1173	2	T31421	C-terminal domain-
41	8	3.2	20	2	S28405	lamtin B receptor -
42	7	3.2	61	2	S39425	protamine P1 - duc
43	7	3.2	114	2	C90219	SU1 ribosomal prot
44	7	3.2	130	2	T20613	hypothetical prote
45	7	3.2	132	2	H70759	hypothetical prote

ALIGNMENTS

RESULT 1
T12483
hypothetical protein DKFP564B0769.1 - human (Fragment)
C/Species: Homo sapiens (man)
C/Date: 23-Jul-1999 #sequence_revision 23-Jul-1999 #text_change 23-Jul-1999
C/Accession: T12483
R/Annotate, W.; Wilkner, U.; Mewes, H.W.; Gassenhuber, U.; Wiemann, S.
submitted to the Protein Sequence Database, June 1999
A/Reference number: Z17527
A/Accession: T12483
A/Status: preliminary
A/Molecule type: mRNA
A/Residues: 1-299 <ANS>
A/Cross-references: EMBL:AL080186
A/Experimental source: fetal brain; clone DKFP564B0769
C/Genetics:
A/Note: DKFP564B0769.1

Query Match
Best Local Similarity 100.0%; Score 9; DB 2; Length 299;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 125 ERRRSRSRS 133
Db 109 ERRRSRSRS 117

RESULT 2
T06671
hypothetical protein T17F15.10 - Arabidopsis thaliana
C/Species: Arabidopsis thaliana (mouse-ear cress)
C/Date: 22-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 22-Oct-1999
C/Accession: T06671
R/Querier, F.; Choisme, N.; Robert, C.; Brottier, P.; Mincker, P.; Catolico, L.; Artigau
submitted to the Protein Sequence Database, April 1999
A/Reference number: Z15793
A/Accession: T06671
A/Molecule type: DNA
A/Residues: 1-347 <QUE>
A/Cross-references: EMBL:AL049658; GSPDB:GN00061; ATSP:T17F15.10
C/Genetics:
A/Experimental source: cultivar Columbia; BAC clone T17F15
A/Gene: ANSP:T17F15.10
A/Map position: 3
A/Introns: 188/3; 206/1; 235/1; 251/2

Query Match
Best Local Similarity 100.0%; Score 9; DB 2; Length 347;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 125 ERRRSRSRS 133
|||||||

Db 129 ERRRSRSRS 137

RESULT 3

155595
C:Species: Homo sapiens (man)
C:Date: 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change 24-Sep-1999
C:Accession: 155595
R:mai, H.; Chan, E.K.; Kiyosawa, K.; Fu, X.D.; Tan, E.M.
J. Clin. Invest. 92, 2419-2426, 1993
A:Title: Novel nuclear autoantigen with splicing factor motifs identified with antibody
A:Reference number: 155595; MUID:94043761; PMID:8227358
A:Accession: 155595
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-524 <RES>
A:Cross-references: GB:LI0910; NID:9405191; PIDN:AAA16346.1; PID:9405192
C:Genetics:
A:Gene: CCJ.3
C:Superfamily: unassigned ribonucleoprotein repeat-containing proteins; ribonucleoprotein
F:154-220/Domains: ribonucleoprotein repeat homology <RML>
F:251-318/Domains: ribonucleoprotein repeat homology <RML>

Query Match 4.2%; Score 9; DB 2; Length 524;
Best Local Similarity 100.0%; Pred. No. 0.51;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 125 ERRRSRSRS 133
Db 78 ERRRSRSRS 86

RESULT 4

VHNVAC
nucleocapsid protein - Autographa californica nuclear polyhedrosis virus
N:Alternate names: arginine-rich 6.9K protein
C:Species: Autographa californica nuclear polyhedrosis virus, AcMNPV
C:Date: 31-Mar-1988 #sequence_revision 31-Mar-1988 #text_change 24-Nov-1999
C:Accession: A26593; E72862
R:Wilson, M.E.; Mainprize, T.H.; Friesen, P.D.; Miller, L.K.
J. Virol. 61, 661-666, 1987
A:Title: Location, transcription, and sequence of a baculovirus gene encoding a small an
A:Reference number: A26593; MUID:87112972; PMID:3543402
A:Accession: A26593
A:Molecule type: mRNA
A:Residues: 1-55 <MTL>
A:Cross-references: GB:M15370; NID:9332483; PIDN:AAA46745.1; PID:9332484
R:Ayres, M.D.; Howard, S.C.; Kuzio, J.; Lopez-Ferber, M.; Possee, R.D.
Virology 202, 586-605, 1994
A:Title: The complete DNA sequence of Autographa californica nuclear polyhedrosis virus.
A:Reference number: A72850; MUID:94303173; PMID:8030224
A:Accession: E72862
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-55 <ATR>
A:Cross-references: GB:LI2858; NID:9510708; PIDN:AAA66730.1; PID:9559169
C:Genetics:
A:Gene: Ac-pe.9
C:Superfamily: Autographa californica nuclear polyhedrosis virus nucleocapsid protein
C:Keywords: nucleocapsid

Query Match 3.7%; Score 8; DB 1; Length 55;
Best Local Similarity 100.0%; Pred. No. 0.87;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 116 YRRSRRS 123
Db 36 YRRSRRS 43

RESULT 5
S34045

protamine - North American opossum

C:Species: Didelphis virginiana, Didelphis marsupialis virginiana (North American opossum
C:Date: 06-Jan-1995 #sequence_revision 06-Jan-1995 #text_change 23-Jul-1999
C:Accession: S34045
R:Winkler, R.J.; Nishikawa, S.; Connor, W.; Dixon, G.H.
Eur. J. Biochem. 215, 63-72, 1993
A:Title: Characterization of a marsupial sperm protamine gene and its transcripts from ti
A:Reference number: S34045; MUID:93345500; PMID:8344286
A:Accession: S34045
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-58 <MIN>
A:Cross-references: EMBL:X74044; NID:9407062; PIDN:CAA52193.1; PID:9407063
C:Superfamily: sperm histone
C:Keywords: DNA binding; nucleus

Query Match 3.7%; Score 8; DB 2; Length 58;
Best Local Similarity 100.0%; Pred. No. 0.91;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 126 RRRSRRS 133
Db 5 RRRSRRS 12

RESULT 6

VHNVAC
nucleocapsid protein DNA binding P6.9 orf100 - Bombyx mori nuclear polyhedrosis virus (1;
N:Alternate names: DNA-binding protein
C:Species: Bombyx mori nuclear polyhedrosis virus, BMSNPV
A:Variety: isolate T3
C:Date: 31-Mar-1992 #sequence_revision 31-Mar-1992 #text_change 11-May-2000
C:Accession: A38554; T41841
R:Maeda, S.; Kamita, S.G.; Kataoka, H.
Virology 180, 807-810, 1991
A:Title: The basic DNA-binding protein of Bombyx mori nuclear polyhedrosis virus: the ex
A:Reference number: A38554; MUID:91112001; PMID:1703373
A:Accession: A38554
A:Molecule type: DNA
A:Residues: 1-65 <MAE>
A:Cross-references: EMBL:M63416; NID:9332494; PIDN:AAA46691.1; PID:9332495
R:Gomi, S.; Majima, K.; Maeda, S.
J. Gen. Virol. 80, 1333-1337, 1999
A:Title: Sequence analysis of the genome of Bombyx mori nucleopolyhedrovirus.
A:Reference number: Z22020; MUID:99281911; PMID:10355780
A:Accession: T41841
A:Status: translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-65 <RAM>
A:Cross-references: EMBL:U33180; PIDN:AA63770.1
A:Experimental source: isolate T3
C:Superfamily: Autographa californica nuclear polyhedrosis virus nucleocapsid protein
C:Keywords: DNA binding; late protein; nucleocapsid

Query Match 3.7%; Score 8; DB 1; Length 65;
Best Local Similarity 100.0%; Pred. No. 1;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 116 YRRSRRS 123
Db 46 YRRSRRS 53

RESULT 7

D35846
transformer-2 sex-determining protein, form D - fruit fly (Drosophila melanogaster)
C:Species: Drosophila melanogaster
C:Date: 23-Oct-1990 #sequence_revision 23-Oct-1990 #text_change 23-Jul-1999
C:Accession: D35846
R:Mattox, W.; Palmer, M.J.; Baker, B.S.
Genes Dev. 4, 789-805, 1990
A:Title: Alternative splicing of the sex determination gene transformer-2 is sex-specific
A:Reference number: A35846; MUID:90337316; PMID:2116360

A:Accession: D35846
 A:Molecule type: DNA
 A:Residues: 1-136 <MAT>
 A:Cross-references: GB:M30939; NID:9158648; PIDN:AAA828957.1; PID:9158652
 C:Genetics:
 A:Gene: tra-2
 A:Cross-references: FlyBase:FBgn0003742
 C:Superfamily: transformer-2 sex-determining protein; ribonucleoprotein repeat homology
 C:Keywords: alternative splicing

Query Match 3.7%; Score 8; DB 2; Length 136;
 Best Local Similarity 100.0%; Pred. No. 1.8;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 126 RRRSRSS 133
 Db 56 RRRSRSS 63

RESULT 8
 I38191
 nucleic acid binding protein - human (fragment)
 C:Species: Homo sapiens (man)
 C>Date: 29-May-1998 #sequence_revision 29-May-1998 #text_change 05-Nov-1999
 C:Accession: I38191
 R:Nakamura, Y.; Morinuchi, R.; Nakayama, D.; Yamashita, I.; Higashiyama, Y.; Yamamoto, T.
 J. Gen. Virol. 75, 2625-2633, 1994
 A:Title: Altered expression of a novel cellular gene as a consequence of integration of
 A:Reference number: I38191; MUID:95016514; PMID:7931148
 A:Accession: I38191
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: mRNA
 A:Residues: 1-163 <RES>
 A:Cross-references: EMBL:X76302; NID:9431952; PIDN:CAA53949.1; PID:9431953
 C:Genetics:
 A:Gene: RY-1
 A:Introns: 124/3; 146/2

Query Match 3.7%; Score 8; DB 2; Length 163;
 Best Local Similarity 100.0%; Pred. No. 2.1;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 126 RRRSRSS 133
 Db 46 RRRSRSS 53

RESULT 9
 T34145
 hypothetical protein C33H5.12 - Caenorhabditis elegans
 C:Species: Caenorhabditis elegans
 C>Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 31-Jan-2000
 C:Accession: T34145
 R:Bradshaw, H.; Stellyes, L.
 submitted to the EMBL Data Library, November 1995
 A:Description: The sequence of C. elegans cosmid C33H5.
 A:Reference number: Z21482
 A:Accession: T34145
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-179 <BRA>
 A:Cross-references: EMBL:U1007; PIDN:AAA82270.1; CESP:C33H5.12
 A:Experimental source: strain Bristol N2
 C:Genetics:
 A:Gene: CESP:C33H5.12
 A:Introns: 18/3; 62/2; 98/2; 149/2
 C:Superfamily: unassigned ribonucleoprotein repeat-containing proteins; ribonucleoprotein

Query Match 3.7%; Score 8; DB 2; Length 179;
 Best Local Similarity 100.0%; Pred. No. 2.3;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 60 DVDAEDA 67

Db 48 DVDAEDA 55
 RESULT 10
 E84638
 probable RSP22 splicing factor [imported] - Arabidopsis thaliana
 C:Species: Arabidopsis thaliana (mouse-ear cress)
 C>Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 14-Sep-2001
 C:Accession: E84638
 R:Jin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;
 M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; Vanden, S.E.; Umayam, L.; Tallon, L.
 eus, D.; Nerman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter,
 Nature 402, 761-768, 1999
 A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
 A:Reference number: A84420; MUID:20083487; PMID:10617197
 A:Accession: E84638
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-196 <STO>
 A:Cross-references: GB:A802093; NID:94572679; PIDN:AAD23894.1; GSPDB:GN00139
 C:Genetics:
 A:Gene: At2g24590
 A:Map position: 2
 C:Superfamily: glycine-rich RNA-binding protein; ribonucleoprotein repeat homology

Query Match 3.7%; Score 8; DB 2; Length 196;
 Best Local Similarity 100.0%; Pred. No. 2.5;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 126 RRRSRSS 133
 Db 120 RRRSRSS 127

RESULT 11
 S26047
 transformer sex-determining protein - fruit fly (Drosophila virilis)
 C:Species: Drosophila virilis
 C>Date: 20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change 17-Mar-1999
 C:Accession: S26047
 R:O'Neil, M.T.; Belote, J.M.
 Genetics 131, 113-128, 1992
 A:Title: Interspecific comparison of the transformer gene of drosophila reveals an unsubst
 A:Reference number: S26044; MUID:92275311; PMID:1552233
 A:Accession: S26047
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-199 <ONE>
 A:Cross-references: EMBL:X66528; NID:99206; PID:99207
 C:Genetics:
 A:Gene: FlyBase:Vtr/tra
 A:Cross-references: FlyBase:FBgn0013147
 A:Introns: 11/3; 138/3

Query Match 3.7%; Score 8; DB 2; Length 199;
 Best Local Similarity 100.0%; Pred. No. 2.5;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 126 RRRSRSS 133
 Db 104 RRRSRSS 111

RESULT 12
 S26045
 sex-determining protein transformer - fruit fly (Drosophila hydei)
 C:Species: Drosophila hydei
 C>Date: 19-Jul-1996 #sequence_revision 19-Jul-1996 #text_change 21-Jul-2000
 C:Accession: S26045
 R:O'Neill, M.T.; Belote, J.M.
 Genetics 131, 113-128, 1992
 A:Title: Interspecific comparison of the transformer gene of drosophila reveals an unsubst

A:Reference number: S26044; MUID:92275311; PMID:1592233
A:Accession: S26045
A:Molecule type: DNA
A:Residues: 1-201 <ONE>
A:Cross-references: EMBL:X66931; NID:97457; PIDD:CAA47365.1; PID:97458
A:Note: the authors did not translate the codon for residue 111
C:Genetics:
A:Gene: tra
A:Cross-references: FlyBase:FBgn0012404
A:introns: 11/3; 134/3

Query Match 3.7%; Score 8; DB 2; Length 201;
Best Local Similarity 100.0%; Pred. No. 2.5;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 126 RRRSRSRS 133
DB 102 RRRSRSRS 109

RESULT 13
PR264 protein - chicken
C:Species: Gallus gallus (chicken)
C>Date: 26-May-1994 #sequence_revision 26-May-1994 #text_change 24-Sep-1999
C:Accession: B42701; S17327
R:Veillard, M.; Sureau, A.; Soret, J.; Martinerie, C.; Perbal, B.
Proc. Natl. Acad. Sci. U.S.A. 89, 2511-2515, 1992
A:Title: A potential splicing factor is encoded by the opposite strand of the trans-spl
A:Reference number: A42701; MUID:92212859; PMID:1557353
A:Accession: B42701
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-221 <VEL>
A:Cross-references: EMBL:X62446; NID:963751; PIDD:CAA4306.1; PID:963752
C:Superfamily: unassigned ribonucleoprotein repeat-containing proteins; ribonucleoprotein
F:15-82/Domain: ribonucleoprotein repeat homology <RM2>

Query Match 3.7%; Score 8; DB 2; Length 221;
Best Local Similarity 100.0%; Pred. No. 2.7;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 126 RRRSRSRS 133
DB 125 RRRSRSRS 132

RESULT 14
AA2701
splicing factor SFRS2 - human
N:Alternate names: PR264/SC35; splicing factor arginine/serine rich 2 (SFRS2)
C:Species: Homo sapiens (man)
C>Date: 26-May-1994 #sequence_revision 26-May-1994 #text_change 17-Nov-2000
C:Accession: I38128; A42701; A42634; I59216; S17328
R:Sureau, A.; Perbal, B.
Proc. Natl. Acad. Sci. U.S.A. 91, 932-936, 1994
A:Title: Several mRNAs with variable 3' untranslated regions and different stability enc
A:Reference number: A36992; MUID:94134745; PMID:8502870
A:Accession: I38128
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-221 <SUR>
A:Cross-references: EMBL:X75755; NID:9455418; PIDD:CAA53383.1; PID:9455419
R:Veillard, M.; Sureau, A.; Soret, J.; Martinerie, C.; Perbal, B.
Proc. Natl. Acad. Sci. U.S.A. 89, 2511-2515, 1992
A:Title: A potential splicing factor is encoded by the opposite strand of the trans-spl
A:Reference number: A42701; MUID:92212859; PMID:1557353
A:Accession: A42701
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-221 <VEL>
A:Cross-references: EMBL:X62447; NID:935596; PIDD:CAA44307.1; PID:935597
R:Pu, X.D.; Mantatis, T.

Science 256, 535-538, 1992
A:Title: Isolation of a complementary DNA that encodes the mammalian splicing factor SC3;
A:Reference number: A42634; MUID:92237694; PMID:1373910
A:Accession: A42634
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-37, 'R', 39-221 <FUA>
A:Cross-references: GB:M90104; NID:9337925; PIDD:AAA60306.1; PID:9337926
R:Sureau, A.; Soret, J.; Veillard, M.; Crochet, J.; Perbal, B.
Proc. Natl. Acad. Sci. U.S.A. 89, 11683-11687, 1992
A:Title: The PR264/c-myb connection: expression of a splicing factor modulated by a nucle
A:Reference number: I59216; MUID:93101590; PMID:1465383
A:Accession: I59216
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-34 <RES>
A:Cross-references: GB:L03693; NID:9292398; PIDD:AAA60162.1; PID:9553620
C:Genetics:
A:Gene: GDB:SFRS2; SC-35
A:Cross-references: GDB:I32412; OMIM:600813
A:introns: 121/2
C:Superfamily: unassigned ribonucleoprotein repeat-containing proteins; ribonucleoprotein
F:15-82/Domain: ribonucleoprotein repeat homology <RM1>

Query Match 3.7%; Score 8; DB 2; Length 221;
Best Local Similarity 100.0%; Pred. No. 2.7;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 126 RRRSRSRS 133
DB 125 RRRSRSRS 132

RESULT 15
A57198
splicing factor, arginine/serine-rich 7 - human
N:Alternate names: splicing factor 9G8
C:Species: Homo sapiens (man)
C>Date: 24-May-1996 #sequence_revision 24-May-1996 #text_change 24-Sep-1999
C:Accession: A57198; S46319
R:Popelar, M.; Cavaloc, Y.; Mattei, M.G.; Gattoni, R.; Stevenin, J.
J. Biol. Chem. 270, 17830-17835, 1995
A:Title: The gene encoding human splicing factor 9G8. Structure, chromosome localization,
A:Reference number: A57198; MUID:9535374; PMID:7629084
A:Accession: A57198
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-238 <RES>
A:Cross-references: GB:L41887; NID:950423; PIDD:AAA88098.1; PID:950424
R:Cavaloc, Y.; Popelar, M.; Fuchs, J.P.; Gattoni, R.; Stevenin, J.
EMBO J. 13, 2639-2649, 1994
A:Title: Characterization and cloning of the human splicing factor 9G8: a novel 35 kDa f
A:Reference number: S46319; MUID:94283389; PMID:8013463
A:Accession: S46319
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-238 <CAV>
A:Cross-references: EMBL:L22253; NID:9506401; PIDD:AAA35495.1; PID:9506402
C:Genetics:
A:Gene: GDB:SFRS7; 9G8
A:Cross-references: GDB:378350; OMIM:600572
A:Map position: 2p22-2p21
A:introns: 10/1; 70/2; 129/2; 154/2; 191/2; 209/2; 221/2
C:Superfamily: unassigned ribonucleoprotein repeat-containing proteins; ribonucleoprotein
F:12-74/Domain: ribonucleoprotein repeat homology <RM3>

Query Match 3.7%; Score 8; DB 2; Length 238;
Best Local Similarity 100.0%; Pred. No. 2.9;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 126 RRRSRSRS 133
DB 125 RRRSRSRS 132

Thu Jul 29 11:52:12 2004

us-09-787-491b-1.rpr

Page 5

Search completed: July 22, 2004, 17:20:59
Job time : 16 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 22, 2004, 17:13:21 ; Search time 13 Seconds

(without alignments)
865.166 Million cell updates/sec

Title: US-09-787-491B-1

Perfect score: 216
Sequence: 1 MSRTLRPPNPTSLFVRNVD.....PKGGMKYLQYECYTNILTLV 216

Scoring table:

Gapop 60.0 , Gapext 60.0

Searched: 141681 seqs, 52070155 residues

Word size: 0

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database: SwissProt_42.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	9	4.2	208	1	YK69 CAEEL
2	9	4.2	530	1	RNP2_HUMAN
3	9	4.2	530	1	RNP2_MOUSE
4	8	3.7	54	1	BVCP_NPVC
5	8	3.7	57	1	HSP1_DIDMA
6	8	3.7	58	1	BVCP_GVCL
7	8	3.7	60	1	HSP1_DASVI
8	8	3.7	61	1	HSP1_SARHA
9	8	3.7	64	1	BVCP_NPYBM
10	8	3.7	199	1	TRSF_DROVI
11	8	3.7	201	1	TRSF_DROHY
12	8	3.7	221	1	SPR2_CHICK
13	8	3.7	221	1	SPR2_HUMAN
14	8	3.7	221	1	SPR2_MOUSE
15	8	3.7	238	1	SPR7_HUMAN
16	8	3.7	252	1	VEHE_NPVAC
17	8	3.7	264	1	TRA2_DROME
18	8	3.7	282	1	TR2B_HUMAN
19	8	3.7	288	1	TR2B_HUMAN
20	8	3.7	344	1	SPR6_HUMAN
21	8	3.7	378	1	RUI7_MOUSE
22	8	3.7	437	1	RUI7_HUMAN
23	8	3.7	448	1	RUI7_DROME
24	8	3.7	471	1	RUI7_XENIA
25	8	3.7	488	1	VE2_HPV49
26	8	3.7	653	1	SPRG_MOUSE
27	8	3.7	659	1	SPRG_MOUSE
28	8	3.7	887	1	A4_DROME
29	8	3.7	919	1	AMPY_STRLI
30	7	3.2	60	1	HSP1_ORNAN
31	7	3.2	114	1	RS17_SULO
32	7	3.2	126	1	RS12_CANTP
33	7	3.2	132	1	YK10_MYCTU

34	7	3.2	148	1	H1L MYTCA	P22974 mytilus cal
35	7	3.2	196	1	SPR2 CAEEL	O09511 caenorhabdi
36	7	3.2	208	1	YKX2 CAEEL	Q10021 caenorhabdi
37	7	3.2	240	1	YD37_MYCTU	Q10647 mycobacteri
38	7	3.2	244	1	YK51 CAEEL	P34433 caenorhabdi
39	7	3.2	303	1	SPR1_ARATH	O22315 arabidopsis
40	7	3.2	306	1	UL31 HSV1	P10215 herpes simp
41	7	3.2	309	1	CYP9 CAEEL	O09637 caenorhabdi
42	7	3.2	325	1	LDH_TACCA	P00343 lactobacilli
43	7	3.2	333	1	SW_TASMO	P57956 pasteurella
44	7	3.2	339	1	SRRI_HUMAN	Q9uh36 homo sapien
45	7	3.2	375	1	SR55_DROME	P26686 drosophila

ALIGNMENTS

RESULT 1
YK69 CAEEL STANDARD; PRT; 208 AA.
ID AC P34347;
DT 01-FEB-1994 (Rel. 28, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Hypothetical protein C29E4.9 in chromosome III.
GN C29E4.9.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Pelodierinae; Caenorhabditis.
OK NCBI TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RX MEDLINE=94150718; PubMed=7906398;

Wilson R., Rinscough R., Anderson K., Baynes C., Berks M.,
Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,
Craxton M., Dear S., Du Z., Durbin R., Favello A., Fraser A.,
Fulton L., Gardner A., Green P., Hawkins T., Hillier L., Jier M.,
Johnston L., Jones M., Kershaw J., Kirsten J., Laister N.,
Latrelle P., Lightning U., Lloyd C., Mortimore B., O'Callaghan M.,
Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Showkeen R.,
Stine M., Smaildon N., Smith A., Smith M., Sonhammer E., Staden R.,
Sulston J., Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K.,
Waterson R., Watson A., Weinstock L., Wilkinson-Sproat J.,
Wohlman P.;
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
elegans.";
RT Nature 368:32-38 (1994).
RL [2]
RN REVISIONS.
RP Waterston R.;
RA Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; I23651; AAA27956.2; -
DR PIR; S44765; S44765.
DR WormBep; C29E4.9; CE29199.
KW Hypothetical protein.
SQ SEQUENCE 208 AA; 23729 MW; FDA57F6609EEB32 CRC64;

Query Match 4.2%; Score 9; DB 1; Length 208;
Best local Similarity 100.0%; Pred. No. 0.07;

Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 125 ERRRRSRS 133
|||||||

DB 153 ERRSRSSRS 161

RESULT 2
RNP2_HUMAN STANDARD; PRT; 530 AA.

AC Q14496; Q14496; (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE RNA-binding region containing protein 2 (Hepatocellular carcinoma protein 1) (splicing factor HCC1).
GN RNP2 OR HCC1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORMS 1 AND 2), AND CHARACTERIZATION.
RC TISSUE=Liver;
RX MEDLINE=94043761; PubMed=8227358;
RA Imai H., Chan E.K.L., Kiyosawa K., Fu X.-D., Tan E.M.;
RA "Novel nuclear autoantigen with splicing factor motifs identified with
RT antibody from hepatocellular carcinoma.";
RL J. Clin. Invest. 92:2419-2426(1993).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=21638749; PubMed=11780052;
RA Deloukas P., Matthews L.H., Ashurst J., Burton J., Gilbert J.G.R.,
RA Jones M., Stavrides G., Almeida J.P., Babbage A.K., Bagunley C.L.,
RA Bailey J., Barlow K.F., Bates K.N., Beard L.M., Beare D.M.,
RA Beasley O.P., Bird C.P., Blakey S.E., Bridgeman A.M., Brown A.J.,
RA Buckley D., Burrill W.D., Butler A.P., Carder C., Carter N.P.,
RA Chapman J.C., Clamp M., Clark G., Clark L.N., Clark S.Y., Clee C.M.,
RA Clegg S., Cobley V.E., Collier R.E., Connor R.E., Corby N.R.,
RA Coulson A.G., Coville G.J., Deadman R., Dhani P.D., Dunn M.,
RA Ellington A.G., Frankland J.A., Fraser A., French L., Garner P.,
RA Graham D.V., Griffiths C., Griffiths M.N.D., Gilliam R., Hall R.E.,
RA Hammond S., Hartley J.L., Heath P.D., Ho S., Holden J.D., Howden P.J.,
RA Huckle E., Hunt A.R., Hunt S.E., Jekosch K., Johnson C.M., Johnson D.,
RA Kay M.P., Kimberley A.M., King A., Knights A., Laird G.K., Lawlor S.,
RA Levasaiho M.H., Levertha M.A., Lloyd C., Lloyd D.M., Lovell J.D.,
RA Marsh V.L., Martin S.L., McCormachie L.J., McElay K., McMurtry A.A.,
RA Milne S.A., Mistry D., Moore M.J.F., Mullikin J.C., Nickerson T.,
RA Oliver K., Parker A., Patel R., Pearce T.A.V., Peck A.I.,
RA Phillimore B.J.C.T., Prathalingam S.R., Plumb R.W., Ramsay H.,
RA Rice C.M., Ross W.T., Scott C.E., Sehra H.K., Showkeen R., Sims S.,
RA Skuce C.D., Smith M.L., Soderlund C., Steward C.A., Sulston J.E.,
RA Swann R.M., Sycamore N., Taylor R., Tee L., Thomas D.W., Thorpe A.,
RA Tracey A., Tromans A.C., Vaudin M., Wall M., Wallis J.M., Williams S.A.,
RA Whitehead S.L., Whitaker P., Willey D.L., Williams L., Williams D.R., Beck S.,
RA Wilming L., Wray P.W., Hubbard T., Dudbin R.M., Bentley D.R., Beck S.,
RA Rogers J.;
RT "The DNA sequence and comparative analysis of human chromosome 20.";
RL Nature 414:865-871 (2001).

-1- FUNCTION: Transcriptional coactivator for steroid nuclear receptors ESRI/ER-alpha and ESR2/ER-beta, and JUN/AP-1 (By similarity). May be involved in pre-mRNA splicing process.
-1- SUBUNIT: Interacts with NCOA6 (By similarity).
-1- SUBCELLULAR LOCATION: Nuclear; concentrated in nuclear speckles. Colocalizes with the core spliceosomal snRNP proteins.
-1- ALTERNATIVE PRODUCTS:
Event=Alternative splicing; Named isoforms=2;
Name=1; Synonyms=HCC1.4;
IsoId=Q14498-1; Sequence=Displayed;
Name=2; Synonyms=HCC1.3;
IsoId=Q14498-2; Sequence=VSP_005820;
-1- TISSUE SPECIFICITY: Widely expressed. Highly expressed in pancreas, skeletal muscle, lung and brain. Expressed at intermediate level in kidney, liver and heart.
-1- MISCELLANEOUS: Antibodies against RNP2 are present in sera from a patient with hepatocellular carcinoma who developed several autoantibodies.

CC -1- SIMILARITY: Belongs to the splicing factor SR family.
CC -1- SIMILARITY: Contains 3 RNA recognition motif (RRM) domains.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.ebi.ac.uk/announcements> or send an email to license@ebi.ac.uk).
CC -----
DR EMBL: L10910; AAA16346.1; -;
DR EMBL: L10911; AAA16347.1; -;
DR EMBL: AL357374; CAC11118.1; -;
DR EMBL: AL357374; CAC11119.1; -;
DR PIR: I55595; I55595.
DR HSSP: P1940; ICVT.
DR Genew: HGNC:15923; RNP2.
DR GK: Q14498; -;
DR MIM: 604739; -;
DR InterPro: IPR000504; RNA_rec_mot.
DR Pfam: PF00076; Rrm; 3.
DR SMART: SM00360; RRM; 3.
DR PROSITE: PS50102; RRM; 1;
DR PROSITE: PS00030; RRM_RNP_1; 2.
DR Transcription regulation; Activator; Nuclear protein; RNA-binding;
KM mRNA processing; mRNA splicing; Repeat; Alternative splicing;
KM Polymorphism.
FT DOMAIN 41 90 ARG/SER-RICH (RS DOMAIN).
FT DOMAIN 153 230 RNA-BINDING (RRM) 1.
FT DOMAIN 250 328 RNA-BINDING (RRM) 2.
FT DOMAIN 445 508 RNA-BINDING (RRM) 3.
FT DOMAIN 291 355 ACTIVATING DOMAIN (BY SIMILARITY).
FT DOMAIN 291 406 INTERACTION WITH JUN (BY SIMILARITY).
FT DOMAIN 355 406 INTERACTION WITH ESR1 AND ESR2
(BY SIMILARITY).
FT DOMAIN 406 530 INTERACTION WITH NCOA6 (BY SIMILARITY).
FT DOMAIN 409 416 POLY-ALA.
FT VARSPPLIC 392 397 Missing (in isoform 2).
FT VARIANT 2 2 /FTID=VSP_005820.
FT A -> V (in dbSNP:1803701).
FT /FTID=VAR_015228.
SQ SEQUENCE 530 AA; 59379 MW; 0CC610356D4AA040 CRC64;
Query Match 4.2%; Score 9; DB 1; Length 530;
Best local Similarity 100.0%; Pred. No. 0.17; 0; Indels 0; Gaps 0;
Matches 9; Conservative 0; Mismatches 0;
QY 125 ERRSRSSRS 133
DB 78 ERRSRSSRS 86
RESULT 3
RNP2_MOUSE STANDARD; PRT; 530 AA.
AC O8VH51; O8VH51; (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE RNA-binding region containing protein 2 (Coactivator of activating protein-1 and estrogen receptors) (Coactivator of AP-1 and ERS)
DE (Transcription coactivator CAPER).
GN RNP2 OR CAPER.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORMS 1 AND 2), FUNCTION, AND INTERACTION WITH
RP NCOA6; ESR1; ESR2 AND JUN.
RC TISSUE=Liver;
RX MEDLINE=21638469; PubMed=11704680;

RA Jung D.-J., Na S.-Y., Na D.S., Lee J.W.;
 RT "Molecular cloning and characterization of CAPER, a novel coactivator
 of activating protein-1 and estrogen receptors.";
 RL J. Biol. Chem. 277:1229-1234(2002).
 RN (2)
 RP SEQUENCE FROM N.A. (ISOFORM 3), AND SEQUENCE OF 102-530 FROM N.A.
 RP (ISOFORM 1).
 RC STRAIN=C57BL/6; TISSUE=Mammary fibroblast, and Retina;
 RA MEDLINE=22389257; PubMed=12477932;
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 RA Klausner R.D., Collins F.S., Wagner K.H., Scheffer C.F., Schuler G.D.,
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
 RA Stapleton M., Soares M.B., Bonaldo M.F., Cavaant T.L., Scheetz T.E.,
 RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.J.,
 RA Raha S.S., Loquellano N.A., Peters G.J., Abraham R.D., Mullaly S.J.,
 RA Bosak S.A., McEwen P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
 RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Fahey J., Helton E., Kettelman M., Madan A., Rodriguez A.,
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
 RA Butcherfield V.S.N., Krzywinski M.I., Skalska U., Smallus D.E.,
 RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
 RT "Generation and initial analysis of more than 15,000 full-length
 human and mouse cDNA sequences.";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
 CC -1- FUNCTION: Transcriptional coactivator for steroid nuclear
 receptors ESR1/ER-alpha and ESR2/ER-beta, and JUN/AP-1. May be
 involved in pre-mRNA splicing process.
 CC -1- SUBUNIT: Interacts with NCOA6.
 CC -1- SUBCELLULAR LOCATION: Nuclear (Probable).
 CC -1- ALTERNATIVE PRODUCTS:
 CC Event=Alternative splicing; Named isoforms=3;
 CC Comment=Experimental confirmation may be lacking for some
 isoforms;
 CC Name=1; Synonyms=HCCL.4;
 CC IsoId=Q8VH51.1; Sequence=Displayed;
 CC Name=2; Synonyms=HCCL.3;
 CC IsoId=Q8VH51.2; Sequence=VSP_005822;
 CC Name=3;
 CC IsoId=Q8VH51.3; Sequence=VSP_005821, VSP_005822;
 CC -1- SIMILARITY: Belongs to the splicing factor SR family.
 CC -1- SIMILARITY: Contains 3 RNA recognition motif (RRM) domains.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: AY061882; AA32373.1; -;
 CC EMBL: BC004000; AA04000.1; -;
 CC EMBL: BC030493; AA03049.1; -;
 CC MGI: 2157953; Rnpg2.
 CC GO: GO:0005515; F:Protein binding; IPI.
 CC InterPro: IPR006509; CCL_SF.
 CC InterPro: IPR000504; RNA_rec_mot.
 CC Pfam: PF00076; rrm; 3.
 CC SMART: SM00360; RRM; 2.
 CC TIGRfams: TIGR01622; SF-CCL; 1.
 CC PROSITE: PS00102; RRM; 2.
 CC PROSITE: PS00030; RRM_RNP_1; 2.
 CC Transcription regulation; Activator; Nuclear protein; RNA-binding;
 CC mRNA processing; mRNA splicing; Repeat; Alternative splicing.
 CC FT DOMAIN 41 90
 CC FT DOMAIN 153 230 RNA-BINDING (RRM) 1.
 CC FT DOMAIN 250 328 RNA-BINDING (RRM) 2.
 CC FT DOMAIN 445 508 RNA-BINDING (RRM) 3.

FT DOMAIN 291 355 ACTIVATING DOMAIN.
 FT DOMAIN 291 406 INTERACTION WITH JUN.
 FT DOMAIN 355 406 INTERACTION WITH ESR1 AND ESR2.
 FT DOMAIN 406 530 INTERACTION WITH NCOA6.
 FT DOMAIN 409 416 POLY-ALA.
 FT VASPLIC 1 157 Missing (in isoform 3).
 FT VASPLIC 392 397 Missing (in isoform 2 and isoform 3).
 FT VASPLIC 397 397 /FtId=VSP_005822.
 FT CONFLICT 194 194 R -> K (IN REF. 2).
 FT CONFLICT 208 208 R -> P (IN REF. 2).
 SQ SEQUENCE 530 AA; 59493 MW; 3BB891A6B1A8CDE CRC64;
 Query Match 4.2%; Score 9; DB 1; Length 530;
 Best Local Similarity 100.0%; Pred. No. 0.17;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Oy 125 ERRSRSSRS 133
 Db 78 ERRSRSSRS 86
 RESULT 4
 BVCP NPVAC STANDARD; PRT; 54 AA.
 ID BVCP NPVAC
 AC P06545;
 DT 01-JAN-1988 (Rel. 06, Created)
 DT 01-MAY-1992 (Rel. 22, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE DNA-binding protein (Arginine-rich protein 6.9 kDa) (Basic viral core
 protein) (Nucleocapsid protein).
 GN P6.9.
 OS Autographa californica nuclear polyhedrosis virus (AcMNPV).
 CC Viruses; dsDNA viruses, no RNA stage; Baculoviridae;
 CC Nucleopolyhedrovirus.
 CX NCBI_TaxID=46015;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=IA;
 RX MEDLINE=87112972; PubMed=3543402;
 RA Wilson M.E., Mainprize T.H., Friesen P.D., Miller L.K.;
 RT "Location, transcription, and sequence of a baculovirus gene encoding
 a small arginine-rich polypeptide.";
 RL J. Virol. 61:661-666(1987).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C6;
 RX MEDLINE=94303173; PubMed=8030224;
 RA Ayres M.D., Howard S.C., Kuzio J., Lopez-Ferber M., Possee R.D.;
 RT "The complete DNA sequence of Autographa californica nuclear
 polyhedrosis virus.";
 RL Virology 202:586-605(1994).
 RN [3]
 RP SEQUENCE OF 11-54 FROM N.A.
 RC STRAIN=HR3;
 RX MEDLINE=91134998; PubMed=1994581;
 RA Lu A., Carstens E.B.;
 RT "Nucleotide sequence of a gene essential for viral DNA replication in
 the baculovirus Autographa californica nuclear polyhedrosis virus.";
 RL Virology 181:336-347(1991).
 RN [4]
 RP SEQUENCE OF 1-30 FROM N.A.
 RC STRAIN=HR3;
 RA Lu A., Craig A., Carstens E.B.;
 RT Submitted (JUN-1994) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: Thought to be responsible for DNA condensation during
 packaging of the nucleocapsids.
 CC -1- PTM: Probably phosphorylated in infected cells.
 CC -1- SIMILARITY: STRONG TO ARGININE RICH DNA-BINDING PROTEIN OF OTHER
 BACULOVIRUSES.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -

```

CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M15370; AAA46745.1; -
DR EMBL; M57687; AAA67905.1; -
DR EMBL; L22858; AAA66730.1; -
DR EMBL; U10885; AAB08766.1; -
DR PIR; A26593; VHWAVC.
KW Nucleocapsid; DNA-binding; Late protein; Repeat; Phosphorylation.
FT INT MET 0 0 BY SIMILARITY
FT DOMAIN 4 23 2 X 6 AA REPEATS OF R-R-R-S-S.
FT REPEAT 4 9 1.
FT REPEAT 18 23 2.
SQ SEQUENCE 54 AA; 6753 MW; 4FBB136721F1AC35 CRC64;

Query Match 3.7%; Score 8; DB 1; Length 54;
Best Local Similarity 100.0%; Pred. No. 0.24;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 116 YRRSRRS 123
DB 35 YRRSRRS 42

RESULT 5
HSP1_DIDMA STANDARD; PRT; 57 AA.
ID_HSP1_DIDMA
AC P35305;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Sperm protamine P1.
GN PRM1.
OS Didelphis marsupialis virginiana (North American opossum), and
OS Monodelphis domestica (short-tailed grey opossum).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Metatheria; Didelphimorphia; Didelphidae; Didelphis.
OX NCBI_TaxID=9267, 13616;
RN [1]
RP SEQUENCE FROM N.A.
RC SPECIES=D.marsupialis;
RC MEDLINE=93345500; PubMed=8344286;
RA Winkfein R.J., Nishikawa S., Connor W., Dixon G.H.;
RT "Characterization of a marsupial sperm protamine gene and its
RT transcripts from the North American opossum (Didelphis
RT marsupialis).";
RT Eur. J. Biochem. 215:63-72(1993).
RN [2]
RP SEQUENCE FROM N.A.
RC SPECIES=M.domestica;
RC MEDLINE=95215351; PubMed=7700877;
RA Relief J.D., Krajewski C., Western M., Winkfein R.J., Dixon G.H.;
RT "Molecular phylogeny and evolution of marsupial protamine P1 genes.";
RT Proc. R. Soc. Lond., B, Biol. Sci. 259:7-14(1995).
CC -1- FUNCTION: Protamines substitute for histones in the chromatin of
CC sperm during the haploid phase of spermatogenesis. They compact
CC sperm DNA into a highly condensed, stable and inactive complex.
CC -1- SUBCELLULAR LOCATION: Nuclear.
CC -1- TISSUE SPECIFICITY: Testis.
CC -1- SIMILARITY: Belongs to the protamine P1 family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
EMBL; L17007; AAA02812.1; -

```

```

DR EMBL; X74044; CAA52193.1; -
DR EMBL; L35448; AAA74612.1; -
DR PIR; S34045; S34045.
DR InterPro; IPR00221; Protamine_P1.
DR Pfam; PF00260; Protamine_P1; 1.
DR PROSITE; PS00048; PROTAMINE_P1; 1.
KW Chromosomal protein; Nucleosome core; Spermatogenesis; DNA-binding;
KW Testis; DNA condensation; Nuclear protein.
FT INT MET 0 0 BY SIMILARITY
SQ SEQUENCE 57 AA; 7810 MW; 263715B280214E52 CRC64;

Query Match 3.7%; Score 8; DB 1; Length 57;
Best Local Similarity 100.0%; Pred. No. 0.25;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 126 RRRSRRS 133
DB 4 RRRSRRS 11

RESULT 6
BVC1_GVCL STANDARD; PRT; 58 AA.
ID_BVC1_GVCL
AC P41726;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE DNA-binding protein (Arginine-rich protein 7.3 kDa) (Basic viral core
DE protein) (Nucleocapsid protein).
GN P7.3.
OS Cryptophlebia leucocretia granulosis virus (C1GV) (Cryptophlebia
OS leucocretia granulovirus).
OC Viruses; dsDNA viruses, no RNA stage; Baculoviridae; Granulovirus.
OX NCBI_TaxID=35254;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV3;
RC MEDLINE=94292936; PubMed=8021613;
RA Jehle J.A., Backhaus H.;
RT "Genome organization of the DNA-binding protein gene region of
RT Cryptophlebia leucocretia granulosis virus is closely related to that
RT of nuclear polyhedrosis viruses.";
RL J. Gen. Virol. 75:1815-1820(1994).
CC -1- FUNCTION: Thought to be responsible for DNA condensation during
CC packaging of the nucleocapsids.
CC -1- PTM: Probably phosphorylated in infected cells.
CC -1- SIMILARITY: STRONG TO ARGININE RICH DNA-BINDING PROTEIN OF OTHER
CC BACULOVIRUSES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X77048; CAA54337.1; -
KW Nucleocapsid; DNA-binding; Late protein; Repeat; Phosphorylation.
SQ SEQUENCE 58 AA; 7291 MW; 2793877A5590C2E CRC64;

Query Match 3.7%; Score 8; DB 1; Length 58;
Best Local Similarity 100.0%; Pred. No. 0.26;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 126 RRRSRRS 133
DB 12 RRRSRRS 19

RESULT 7
HSP1_DASV1 STANDARD; PRT; 60 AA.
ID_HSP1_DASV1

```

```

AC P42135; P42133;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Sperm protamine P1.
GN PRM1.
OS Dasyurus viverrinus (Southeastern quoll), and
OS Dasyurus hallucatus (Satanellus/northern quoll).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Metatheria; Dasyuromorphia; Dasyuridae; Dasyurus.
OX NCBI_TaxID=9279, 9280;
RN [1]
RP SEQUENCE FROM N.A.
RC SPECIES=D.viverrinus, and D.hallucatus;
RC TISSUE=Sperm;
RA MEDLINE=95215551; PubMed=7700877;
RT "Retief J.D., Krajewski C., Westernman M., Winkfein R.J., Dixon G.H.;
  "Molecular phylogeny and evolution of marsupial protamine P1 genes.";
  Proc. R. Soc. Lond., B, Biol. Sci. 259:7-14(1995).
RL
CC -1- FUNCTION: Protamines substitute for histones in the chromatin of
  sperm during the haploid phase of spermatogenesis. They compact
  sperm DNA into a highly condensed, stable and inactive complex.
CC -1- SUBCELLULAR LOCATION: Nuclear.
CC -1- TISSUE SPECIFICITY: Testis.
CC -1- SIMILARITY: Belongs to the protamine P1 family.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
  between the Swiss Institute of Bioinformatics and the EMBL outstation -
  the European Bioinformatics Institute. There are no restrictions on its
  use by non-profit institutions as long as its content is in no way
  modified and this statement is not removed. Usage by and for commercial
  entities requires a license agreement (See http://www.isb-sib.ch/announce/
  or send an email to license@isb-sib.ch).
-----
DR EMBL; L35340; AAA74599.1; -
DR InterPro; IPR000221; Protamine P1.
DR PROSITE; PS00048; PROTAMINE_P1; 1.
KW Chromosomal protein; Nucleosome core; Spermatogenesis; DNA-binding;
  Testis; DNA condensation; Nuclear protein.
FT INIT MET 0
SQ SEQUENCE 60 AA; 8246 MW; 615D3D85E7123025 CRC64;

Query Match
Best Local Similarity 100.0%; Score 8; DB 1; Length 60;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 126 RRRSRSS 133
Db 4 RRRSRSS 11

RESULT 8
HSP1_SARHA STANDARD; PRT; 61 AA.
AC P42151;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Sperm protamine P1.
GN PRM1.
OS Sarcophilus harrisii (Tasmanian devil), and
OS Dasyurus maculatus (Tiger quoll).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Metatheria; Dasyuromorphia; Dasyuridae; Sarcophilus.
OX NCBI_TaxID=9305, 9281;
RN [1]
RP SEQUENCE FROM N.A.
RC SPECIES=S.harrisii; TISSUE=Sperm;
RA MEDLINE=95215551; PubMed=7700877;
RT "Retief J.D., Krajewski C., Westernman M., Winkfein R.J., Dixon G.H.;
  "Molecular phylogeny and evolution of marsupial protamine P1 genes.";
  Proc. R. Soc. Lond., B, Biol. Sci. 259:7-14(1995).

```

```

RN [2]
RP SEQUENCE FROM N.A.
RC SPECIES=D.maculatus;
RA Krajewski C., Young J., Buckley L., Woolley P.A., Westernman M.;
RT "Reconstructing the taxonomic radiation of dasyurine marsupials with
  cytochrome b, 12S rRNA, and protamine P1 gene trees.";
  J. Mammal. Evol. 4:217-236(1997).
RL
CC -1- FUNCTION: Protamines substitute for histones in the chromatin of
  sperm during the haploid phase of spermatogenesis. They compact
  sperm DNA into a highly condensed, stable and inactive complex.
CC -1- SUBCELLULAR LOCATION: Nuclear.
CC -1- TISSUE SPECIFICITY: Testis.
CC -1- SIMILARITY: Belongs to the protamine P1 family.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
  between the Swiss Institute of Bioinformatics and the EMBL outstation -
  the European Bioinformatics Institute. There are no restrictions on its
  use by non-profit institutions as long as its content is in no way
  modified and this statement is not removed. Usage by and for commercial
  entities requires a license agreement (See http://www.isb-sib.ch/announce/
  or send an email to license@isb-sib.ch).
-----
DR EMBL; L35324; AAA74608.1; -
DR EMBL; AF010276; AAB69306.1; -
DR InterPro; IPR000221; Protamine P1.
DR PROSITE; PS00048; PROTAMINE_P1; 1.
KW Chromosomal protein; Nucleosome core; Spermatogenesis; DNA-binding;
  Testis; DNA condensation; Nuclear protein.
FT INIT MET 0
SQ SEQUENCE 61 AA; 8410 MW; 4A215D3D85E71230 CRC64;

Query Match
Best Local Similarity 100.0%; Score 8; DB 1; Length 61;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 126 RRRSRSS 133
Db 4 RRRSRSS 11

RESULT 9
BVCP_NPVEM STANDARD; PRT; 64 AA.
AC P24649;
DT 01-MAR-1992 (Rel. 21, Created)
DT 01-MAR-1992 (Rel. 22, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE DNA-binding protein (Arginine-rich protein 8.1 kDa) (Basic viral core
  protein) (Nucleocapsid protein).
OS Bombyx mori nuclear polyhedrosis virus (BmNPV).
OC Viruses; dsDNA viruses, no RNA stage; Baculoviridae;
OC Nucleopolyhedrovirus.
OX NCBI_TaxID=10458;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 1-45.
RC STRAIN=T3;
RA MEDLINE=9112001; PubMed=1703373;
RA Maeda S., Kamita S.G., Kataoka H.;
RT "The basic DNA-binding protein of Bombyx mori nuclear polyhedrosis
  virus: the existence of an additional arginine repeat.";
  Virology 180:807-810(1991).
RL [2]
RP SEQUENCE FROM N.A.
RC STRAIN=T3;
RA MEDLINE=99281911; PubMed=10355780;
RA Gomi S., Majima K., Maeda S.;
RT "Sequence analysis of the genome of Bombyx mori
  nucleopolyhedrovirus.";
  J. Gen. Virol. 80:1323-1337(1999).
RL
CC -1- FUNCTION: Thought to be responsible for DNA condensation during
  packaging of the nucleocapsids.
CC -1- PTM: Probably phosphorylated in infected cells.
CC -1- SIMILARITY: STRONG TO ARGININE RICH DNA-BINDING PROTEIN OF OTHER

```

CC BACULOVIRUSES.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to license@isb-sib.ch).

CC -----

CC EMBL: M63416; AAA46691.1; -.

CC DR EMBL: 133180; AAC63770.1; -.

CC DR PIR: A38554; VHWYBM.

CC Nucleocapsid; DNA-binding; Late protein; Repeat; Phosphorylation.

CC FT INT MET 0 0 3 X 6 AA REPEATS OF R-R-R-R-S-S.

CC FT REPEAT 4 33 1.

CC FT REPEAT 4 9 2.

CC FT REPEAT 18 23 3.

CC FT REPEAT 28 33 3.

CC SEQUENCE 64 AA; 7947 MW; 0F1BD3BD1C68B9AE CRC64;

Query Match 3.7%; Score 8; DB 1; Length 64;

Best Local Similarity 100.0%; Pred. No. 0.28;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 116 YRRSRSSRS 123

DB 45 YRRSRSSRS 52

RESULT 10

TRSF_DROVI STANDARD; PRT; 199 AA.

ID TRSF_DROVI STANDARD; PRT; 199 AA.

AC Q24761;

DT 01-NOV-1997 (Rel. 35, Created)

DT 01-NOV-1997 (Rel. 35, Last sequence update)

DT 28-FEB-2003 (Rel. 41, Last annotation update)

DE Female-specific transformer protein.

GN Drosophila virilis (Fruit fly).

OC Eukaryota; Metazoa; Arthropoda; Insecta; Pterygota;

OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;

OC Ephydroidea; Drosophilidae; Drosophila.

OK NCBI_TaxID=7244;

RN [1]

RP SEQUENCE FROM N.A., AND FUNCTION.

RX MEDLINE=92275311; PubMed=1592233;

RA O'Neil M.T., Belote J.M.;

RT "Interspecific comparison of the transformer gene of Drosophila

RT reveals an unusually high degree of evolutionary divergence.";

RL Genetics 131:113-128(1992).

-I- FUNCTION: Member of the regulatory pathway controlling female

CC somatic sexual differentiation, regulated by Sxl. Activates

CC dsx female-specific splicing by promoting the formation of a

CC splicing enhancer complex which consists of tra, tra2 and sr

CC proteins.

CC -I- SUBCELLULAR LOCATION: Nuclear; speckled subnuclear compartment (by

CC similarity).

CC -I- DOMAIN: RS domain directs localization of proteins to the speckled

CC subnuclear compartment and the purpose of this localization is to

CC allow colocalization and co-concentration of components of the

CC splicing and splicing regulatory machinery to permit relatively

CC high rates and/or efficiencies of reaction and interaction (by

CC similarity).

CC -I- MISCELLANEOUS: The sexual regulation of tra occurs through a

CC mechanism of sex-specific alternative RNA splicing. The non-sex-

CC specific RNA expressed in males is not translated.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to license@isb-sib.ch).

CC -----

CC EMBL: X66931; CAA47365.1; -.

CC DR PIR: S26045; S26045.

CC DR FlyBase; FBgn0012404; Dnyd\tra.

CC FT Nucleur protein; Sexual differentiation.

CC FT DOMAIN 61 114 ARG/SER-RICH (RS DOMAIN).

CC SEQUENCE 201 AA; 23743 MW; 1EB759B2AC6DB34D CRC64;

Query Match 3.7%; Score 8; DB 1; Length 201;

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to license@isb-sib.ch).

CC -----

CC EMBL: X66528; CAA47141.1; -.

CC DR PIR: S26047; S26047.

CC DR FlyBase; FBgn0013147; Dvir\tra.

CC FT Nucleur protein; Sexual differentiation.

CC FT DOMAIN 56 118 ARG/SER-RICH (RS DOMAIN).

CC SEQUENCE 199 AA; 23809 MW; 23655867BB539733 CRC64;

Query Match 3.7%; Score 8; DB 1; Length 199;

Best Local Similarity 100.0%; Pred. No. 0.8;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 126 RRRSRSSRS 133

DB 104 RRRSRSSRS 111

RESULT 11

TRSF_DROHY STANDARD; PRT; 201 AA.

ID TRSF_DROHY STANDARD; PRT; 201 AA.

AC Q23949;

DT 01-NOV-1997 (Rel. 35, Created)

DT 01-NOV-1997 (Rel. 35, Last sequence update)

DT 28-FEB-2003 (Rel. 41, Last annotation update)

DE Female-specific transformer protein.

GN Drosophila hydei (Fruit fly).

OC Eukaryota; Metazoa; Arthropoda; Insecta; Pterygota;

OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;

OC Ephydroidea; Drosophilidae; Drosophila.

OK NCBI_TaxID=7224;

RN [1]

RP SEQUENCE FROM N.A., AND FUNCTION.

RX MEDLINE=92275311; PubMed=1592233;

RA O'Neil M.T., Belote J.M.;

RT "Interspecific comparison of the transformer gene of Drosophila

RT reveals an unusually high degree of evolutionary divergence.";

RL Genetics 131:113-128(1992).

-I- FUNCTION: Member of the regulatory pathway controlling female

CC somatic sexual differentiation, regulated by Sxl. Activates

CC dsx female-specific splicing by promoting the formation of a

CC splicing enhancer complex which consists of tra, tra2 and sr

CC proteins.

CC -I- SUBCELLULAR LOCATION: Nuclear; speckled subnuclear compartment (by

CC similarity).

CC -I- DOMAIN: RS domain directs localization of proteins to the speckled

CC subnuclear compartment and the purpose of this localization is to

CC allow colocalization and co-concentration of components of the

CC splicing and splicing regulatory machinery to permit relatively

CC high rates and/or efficiencies of reaction and interaction (by

CC similarity).

CC -I- MISCELLANEOUS: The sexual regulation of tra occurs through a

CC mechanism of sex-specific alternative RNA splicing. The non-sex-

CC specific RNA expressed in males is not translated.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to license@isb-sib.ch).

CC -----

CC EMBL: X66931; CAA47365.1; -.

CC DR PIR: S26045; S26045.

CC DR FlyBase; FBgn0012404; Dnyd\tra.

CC FT Nucleur protein; Sexual differentiation.

CC FT DOMAIN 61 114 ARG/SER-RICH (RS DOMAIN).

CC SEQUENCE 201 AA; 23743 MW; 1EB759B2AC6DB34D CRC64;

Query Match 3.7%; Score 8; DB 1; Length 201;

Best Local Similarity 100.0%; Pred. No. 0.8;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 126 RRRSRSS 133
DB 102 RRRSRSS 109

RESULT 12

SPR2_CHICK
ID SPR2_CHICK STANDARD; PRT; 221 AA.
AC P30352;
DT 01-APR-1993 (Rel. 25, Created)
DT 01-APR-1993 (Rel. 25, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Splicing factor, arginine/serine-rich 2 (Splicing factor SC35) (SC-35)
GN SFRS2.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Thymus;
RX MEDLINE=92212859; PubMed=157353;
RA Vellard M., Sureau A., Soret J., Martinerie C., Perbal B.;
RT "A potential splicing factor is encoded by the opposite strand of the
trans-spliced c-myc exon";
RL Proc. Natl. Acad. Sci. U.S.A. 89:2511-2515(1992).
CC -1- FUNCTION: Necessary for the splicing of pre-mRNA. It is required
CC for formation of the earliest ATP-dependent splicing complex and
CC interacts with spliceosomal components bound to both the 5' and 3'
CC splice sites during spliceosome assembly. It also is required for
CC ATP-dependent interactions of both U1 and U2 snRNPs with pre-mRNA.
CC -1- SUBCELLULAR LOCATION: Nuclear.
CC -1- PIM: Extensively phosphorylated on serine residues in the RS
CC domain (by similarity).
CC -1- SIMILARITY: Belongs to the splicing factor SR family.
CC -1- SIMILARITY: Contains 1 RNA recognition motif (RRM) domain.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collabora-
CC tion between the Swiss Institute of Bioinformatics and the EMBL outstation
CC at the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL; X62446; CAA44306.1; -.
DR PIR; B42701; B42701.
DR HSP; P19339; ISXL.
DR InterPro; IPR000504; RNA_rec_mot.
DR Pfam; PF00076; rrm; 1.
DR SMART; SMD00360; RRM; 1.
DR PROSITE; PSS0102; RRM; 1.
DR PROSITE; PSS00030; RRM_RNP_1; 1.
KW Nuclear protein; RNA-binding; mRNA splicing; Phosphorylation.
FT DOMAIN 14 92 RNA-BINDING (RRM).
FT DOMAIN 111 116 GY-RICH (HINGE REGION).
FT DOMAIN 117 221 ARG/SER-RICH (RS DOMAIN).
SQ SEQUENCE 221 AA; 25524 MW; 75A4D8FP9170F1BF CRC64;

Query Match 3.7%; Score 8; DB 1; Length 221;
Best Local Similarity 100.0%; Pred. No. 0.88;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 126 RRRSRSS 133
DB 125 RRRSRSS 132

RESULT 13
SPR2_HUMAN
ID SPR2_HUMAN STANDARD; PRT; 221 AA.
AC O01130;
DT 01-APR-1993 (Rel. 25, Created)
DT 01-APR-1993 (Rel. 25, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Splicing factor, arginine/serine-rich 2 (Splicing factor SC35) (SC-35)
GN SFRS2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 67-84.
RX MEDLINE=92237694; PubMed=1373910;
RA Fu X.-D., Maniatis T.;
RT "Isolation of a complementary DNA that encodes the mammalian splicing
factor SC35";
RL Science 256:535-538(1992).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Bone marrow;
RX MEDLINE=92212859; PubMed=157353;
RA Vellard M., Sureau A., Soret J., Martinerie C., Perbal B.;
RT "A potential splicing factor is encoded by the opposite strand of the
trans-spliced c-myc exon";
RL Proc. Natl. Acad. Sci. U.S.A. 89:2511-2515(1992).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=94134745; PubMed=8302870;
RA Sureau A., Perbal B.;
RT "Several mRNAs with variable 3' untranslated regions and different
stability encode the human PR264/SC35 splicing factor";
RL Proc. Natl. Acad. Sci. U.S.A. 91:932-936(1994).
RN [4]
RP SEQUENCE FROM N.A.
RC TISSUE=Lung;
RX MEDLINE=22388257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Hutcheson S.F., Zeeberg B., Bueltow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Heide F.,
RA Diatchenko L., Marsina K., Farmer A.A., Rubin G.M., Hong L.,
RA Sleptchenko M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein W.J., Ueudin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullany S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulik S.W.,
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahy J., Helton E., Kettner M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakeley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Buttefield Y.S.N., Krzywinski M.I., Skalek U., Smellus D.E.,
RA Scherch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length
human and mouse cDNA sequences";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [5]
RP SEQUENCE OF 1-34 FROM N.A.
RX MEDLINE=93101590; PubMed=1465383;
RA Sureau A., Soret J., Vellard M., Crochet J., Perbal B.;
RT "The PR264/c-myc connection: expression of a splicing factor
modulated by a nuclear protooncogene";
RL Proc. Natl. Acad. Sci. U.S.A. 89:11683-11687(1992).
RN [6]
RP SEQUENCE OF 67-83.
RX MEDLINE=92249775; PubMed=1577227;
RA Zahler A.M., Lane W.S., Stolk J.A., Roth M.B.;
RT "SR proteins: a conserved family of pre-mRNA splicing factors";
RL Genes Dev. 6:837-847(1992).

[7] INTERACTIONS IN SPliceosome ASSEMBLY.
 RA MEDLINE=94084782; PubMed=8261509;
 RA Wu J.Y., Maniatis T.;
 RT "Specific interactions between proteins implicated in splice site
 RT selection and regulated alternative splicing.";
 RL Cell 75:1061-1070(1993).
 RN [8]
 RP SPECIFICITY FOR BETA-GLOBIN MRNA.
 RX MEDLINE=93368668; PubMed=8361546;
 RA Fu X.-D.;
 RT "Specific commitment of different pre-mRNAs to splicing by single SR
 RT proteins.";
 RL Nature 365:82-85(1993).
 RL [9]
 RP BINDING TO U1-70K.
 RX MEDLINE=94187841; PubMed=8139654;
 RA Kohltz J.D., Jamison S.F., Will C.L., Zuo P., Luhrmann R.,
 RA Garcia-Blanco M.A., Manley J.L.;
 RT "Protein-protein interactions and 5'-splice-site recognition in
 RT mammalian mRNA precursors.";
 RL Nature 368:119-124(1994).
 RL [10]
 RP RNA BINDING SPECIFICITY.
 RX MEDLINE=95354672; PubMed=7543047;
 RA Tacke R., Manley J.L.;
 RT "The human splicing factors ASF/SF2 and SC35 possess distinct,
 RT functionally significant RNA binding specificities.";
 RL EMBO J. 14:3540-3551(1995).
 RN [11]
 RP INTERACTION WITH SFRS12.
 RX MEDLINE=20221551; PubMed=10757789;
 RA Barnard D.C., Patton J.G.;
 RT "Identification and characterization of a novel serine-arginine-rich
 RT splicing regulatory protein.";
 RL Mol. Cell. Biol. 20:3049-3057(2000).
 CC -1- FUNCTION: Necessary for the splicing of pre-mRNA. It is required
 CC for formation of the earliest ATP-dependent splicing complex and
 CC interacts with spliceosomal components bound to both the 5' and 3'
 CC splice sites during spliceosome assembly. It also is required for
 CC ATP-dependent interactions of both U1 and U2 snRNPs with pre-
 CC mRNA. Interacts with other spliceosomal components, via the RS
 CC domains, to form a bridge between the 5' and 3' splice site
 CC binding components, U1 snRNP and U2AF. Binds to
 CC putative-rich RNA sequences, either 5'-AGSAGAGTA-3' (S=C or G) or
 CC 5'-GTTGAGTA-3'. Can bind to beta-globin mRNA and commit it to the
 CC splicing pathway.
 CC -1- SUBUNIT: In vitro, binds SFRS1 (ASF/SF2), SNRP70 and U2AF1 but not
 CC U2AF2. Binds SFRS12.
 CC -1- SUBCELLULAR LOCATION: Nuclear.
 CC -1- PTM: Extensively phosphorylated on serine residues in the RS
 CC domain.
 CC -1- SIMILARITY: Contains 1 RNA recognition motif (RRM) domain.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC at the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; M90104; AAA60306.1; -;
 DR EMBL; X62447; CAA44307.1; -;
 DR EMBL; X75755; CAA53383.1; -;
 DR EMBL; BC000339; AAH00339.1; -;
 DR EMBL; L03693; AAA60162.1; -;
 DR HSSP; P19339; ISXL;
 DR Genew; HGNC:10783; SFRS2.
 DR GK; Q01130; -;
 DR MIM; 600813; -;
 GO; GO:0008248; F:pre-mRNA splicing factor activity; TAS.

DR GO; GO:0006397; P:mRNA processing; TAS.
 DR GO; GO:0006371; P:mRNA splicing; TAS.
 DR InterPro; IPR00504; RNA_rec_mot.
 DR Pfam; PF00076; rrm; 1.
 DR SMART; SM00360; RRM; 1.
 DR PROSITE; PS50102; RRM; 1.
 DR PROSITE; PS00030; RRM_RNP_1; 1.
 KW Nuclear protein; RNA-binding; mRNA splicing; Phosphorylation.
 FT DOMAIN 14 92
 FT DOMAIN 111 116
 FT DOMAIN 117 221
 FT CONFLICT 38 38
 SQ SEQUENCE 221 AA; 25575 MW; 9D1B76BDB5701F5 CRC64;
 Query Match 3.7%; Score 8; DB 1; Length 221;
 Best Local Similarity 100.0%; Pred. No. 0.88;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Oy 126 RRRSRSS 133
 Db 125 RRRSRSS 132
 RESULT 14
 SFR2_MOUSE
 ID SFR2_MOUSE STANDARD; PRT; 221 AA.
 AC Q62093;
 DT 15-DEC-1998 (Rel. 37, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 15-MAR-2004 (Rel. 43, Last annotation update)
 DE Splicing factor, arginine/serine-rich 2 (Splicing factor SC35) (SC-35)
 DE (Splicing component, 35 kDa) (PR264 protein).
 GN SFRS2 OR PR264.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=98447613; PubMed=9774382;
 RA Yang L., Embree L.J., Tsai S., Hickstein D.D.;
 RT "Oncoprotein TLS interacts with serine-arginine proteins involved in
 RT RNA splicing.";
 RL J. Biol. Chem. 273:27761-27764(1998).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=Mammary gland;
 RX MEDLINE=22388257; PubMed=1477932;
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 RA Klausner R.D., Collins F.S., Wagner K.H., Shenmen C.M., Schuler G.D.,
 RA Altschul S.F., Zeeberg B., Burow K.H., Schaefer C.F., Bhat N.K.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
 RA Stadelton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
 RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
 RA Raha S.S., Loggiano N.A., Peters G.J., Abramson R.D., Mullighy S.J.,
 RA Boeck S.A., McEwan P.J., McKernan K.J., Malek J.A., Guarnate P.H.,
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
 RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Fahey U., Helton E., Kettelman M., Madaen A., Rodrigues S., Sanchez A.,
 RA Whiting M., Madaen A., Young A.C., Shevchenko Y., Bouffard G.G.,
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
 RA Butterfield V.S.N., Krzywinski M.I., Skalska U., Smalins D.E.,
 RA Scherch A., Schein J.E., Jones S.J.M., Marra M.A.;
 RT "Generation and initial analysis of more than 15,000 full-length
 RT human and mouse cDNA sequences.";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
 RN [3]
 RP SEQUENCE OF 1-121 FROM N.A.
 RC STRAIN=129/Sv; TISSUE=Liver;
 RA Galliard C., Perbal B.;
 RL Submitted (JUN-1996) to the EMBL/GenBank/DBD databases.

CC -!- FUNCTION: Necessary for the splicing of pre-mRNA. It is required
CC for formation of the earliest ATP-dependent splicing complex and
CC interacts with spliceosomal components bound to both the 5' and
CC 3' splice sites during spliceosome assembly. It also is required
CC for ATP-dependent interactions of both U1 and U2 snRNPs with pre-
CC mRNA (By similarity).
CC -!- SUPPLEMENT: In vitro, binds SF2 (ASF/SF2), SNRP70 and U2AF1 but not
CC U2AF2. Binds SF2 (By similarity).
CC -!- SUBCELLULAR LOCATION: Nuclear.
CC -!- PTM: Extensively phosphorylated on serine residues in the RS
CC domain (By similarity).
CC -!- SIMILARITY: Belongs to the splicing factor SR family.
CC -!- SIMILARITY: Contains 1 RNA recognition motif (RRM) domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF077858; AAC71000.1; -
DR EMBL; BC005493; AAH05493.1; -
DR EMBL; X98511; CA67134.1; -
DR HSSP; P19339; 1SXL.
DR WGI; MGI:98284; Sfrs2.
DR GO; GO:0005681; C:spliceosome complex; IDA.
DR InterPro; IPR000504; C:spliceosome complex; IDA.
DR Pfam; PF00076; rrm; 1.
DR SMART; SM00360; RRM; 1.
DR PROSITE; PS50102; RRM; 1.
DR PROSITE; PS00030; RRM_NRP_1; 1.
DR Nuclear protein; RNA-binding; mRNA splicing; Phosphorylation.
FT DOMAIN 14 92 GLY-RICH (HINCH REGION).
FT DOMAIN 111 116 GLY-RICH (HINCH REGION).
FT DOMAIN 117 221 ARG/SER-RICH (RS DOMAIN).
SQ SEQUENCE 221 AA; 25476 MW; 68121ACAD35714FA CRC64;

Query Match 3.7%; Score 8; DB 1; Length 221;
Best Local Similarity 100.0%; Pred. No. 0.88;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 126 RRRSRSS 133
Db 125 RRRSRSS 132

RESULT 15
SFR7_HUMAN
ID SFR7_HUMAN STANDARD; PRT; 238 AA.
AC Q16629;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Splicing factor, arginine/serine-rich 7 (Splicing factor 968).
GN SFRS7.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 7-24; 30-58 AND 79-87.
RC TISSUE=Placenta;
RC MEDLINE=9428389; PubMed=8013463;
RA Cavaloc Y., Popielarz M., Fuchs J.-P., Gatroni R., Stevenin J.;
RT "Characterization and cloning of the human splicing factor 968: a
RT novel 35 kDa factor of the serine/arginine protein family.";
RL EMBL J. 13:2639-2649(1994).
RU [2]
RP SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.
RC TISSUE=Placenta;
RC MEDLINE=95355374; PubMed=7629084;

RA Popielarz M., Cavaloc Y., Matei M.-G., Gatroni R., Stevenin J.;
RT "The gene encoding human splicing factor 968. Structure, chromosomal
RT localization, and expression of alternatively processed transcripts";
RL J. Biol. Chem. 270:17830-17835(1995).
RX [3]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain, Cervix, Prostate, and Skin;
RC MEDLINE=22388257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marzina K., Farmer A.A., Rubin G.M., Hong L.,
RA Strophein M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Uscin T.B., Toshlyak S., Carrion P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
RA Bosak S.A., McEwen P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butlerfield Y.S.N., Krzywinski M.I., Skalska V., Smallus D.E.,
RA Scherch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length
RT human and mouse cDNA sequences";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16699-16903(2002).
RX [4]
RP INTERACTION WITH CCN1, CDC2L1, AND CDC2L2.
RX MEDLINE=2249644; PubMed=12501247;
RA Hu D., Mayeda A., Trembley J.H., Lahti J.M., Kidd V.J.;
RT "CDK1 complexes promote pre-mRNA splicing";
RL J. Biol. Chem. 278:8623-8629(2003).
CC -!- FUNCTION: Required for pre-mRNA splicing. Can also modulate
CC alternative splicing in vitro.
CC -!- SUBUNIT: Found in large molecular weight complexes containing
CC CCN1 and the p110 isoforms of either CDC2L1 or CDC2L2.
CC -!- SUBCELLULAR LOCATION: Nuclear.
CC -!- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=3;
CC Comment=Isoforms, often lacking the RS domain and differentially
CC expressed in fetal tissues, may be involved in modulation of 968
CC function;
CC Name=1;
CC IsoId=Q16629-1; Sequence=Displayed;
CC Name=2;
CC IsoId=Q16629-2; Sequence=VSP_005872, VSP_005873;
CC Name=3;
CC IsoId=Q16629-3; Sequence=VSP_005874, VSP_005875;
CC -!- TISSUE SPECIFICITY: Brain, liver, kidney and lung.
CC -!- PTM: Extensively phosphorylated on serine residues in the RS
CC domain.
CC -!- SIMILARITY: Belongs to the splicing factor SR family.
CC -!- SIMILARITY: Contains 1 CCHC-type zinc finger.
CC -!- SIMILARITY: Contains 1 RNA recognition motif (RRM) domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; L22253; AAA35495.1; -
DR EMBL; L41887; AAA88098.1; -
DR EMBL; BC000997; AAH00997.1; -
DR EMBL; BC017369; AAH17369.1; -
DR EMBL; BC017908; AAH17908.1; -
DR EMBL; BC022328; AAH22328.1; -
DR PIR; AS7198; AS7198.
DR HSSP; P11940; 1CVJ.

```

DR   Genew: HGNC:10789; SFRS7.
DR   GK; Q16629; -.
DR   MIM; 600572; -.
DR   GO; GO:0005634; C:nucleus; TAS.
DR   GO; GO:0008248; F:pre-mRNA splicing factor activity; TAS.
DR   GO; GO:0006397; P:mRNA processing; TAS.
DR   GO; GO:0006371; P:mRNA splicing; TAS.
DR   InterPro; IPR000504; RNA_rec mot.
DR   InterPro; IPR01878; Znf_CCHC.
DR   Pfam; PF00076; rrm; 1.
DR   Pfam; PF00098; zf-CCHC; 1.
DR   PRINTS; PR00939; C2HCZNFINGER.
DR   SMART; SM00360; RRM; 1.
DR   SMART; SM00343; Znf_C2HC; 1.
DR   PROSITE; PS50102; RRM; 1. 1. FALSE_NEG.
DR   PROSITE; PS00030; RRM_RNP_1;
DR   PROSITE; PS50158; zf_CCHC; 1.
DR   Nuclear protein; RNA-binding; mRNA splicing; Alternative splicing;
DR   Phosphorylation; Repeat; Zinc-finger.
DR   DOMAIN 11 84 RNA-BINDING (RRM).
FT   ZN_FING 104 120 CCHC-TYPE.
FT   DOMAIN 121 238 ARG/SER-RICH (RS DOMAIN).
FT   DOMAIN 153 227 6 X 8 AA REPEATS OF R-R-S-R-S-X-S-X.
FT   REPEAT 153 160 1.
FT   REPEAT 161 168 2.
FT   REPEAT 169 176 3.
FT   REPEAT 177 184 4.
FT   REPEAT 211 218 5 (APPROXIMATE).
FT   REPEAT 219 226 6 (APPROXIMATE).
FT   REPEAT 219 226 SRSRSH -> AENLR (in isoform 2).
FT   VARSPLIC 130 135 /FTId=VSP_005872.
FT   VARSPLIC 136 238 Missing (in isoform 2).
FT   VARSPLIC 130 132 /FTId=VSP_005873.
FT   VARSPLIC 133 238 SRS -> YLF (in isoform 3).
FT   VARSPLIC 133 238 Missing (in isoform 3).
FT   VARSPLIC 133 238 /FTId=VSP_005875.
SQ   SEQUENCE 238 AA; 27366 MW; 49136754D9630853 CRC64;

```

Query Match

Best Local Similarity 3.7%; Score 8; DB 1; Length 238;
 Matches 8; Conservative 100.0%; Pred.No. 0.94;
 Mismatches 0; Indels 0; Gaps 0;

```

QY   126 RRRSRSS 133
DB   125 RRRSRSS 132

```

Search completed: July 22, 2004, 17:19:21
 Job time : 22 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 22, 2004, 17:13:52 ; Search time 39 Seconds

(without alignments)
1747.485 Million cell updates/sec

Title: us-09-787-491b-1

Perfect score: 216
Sequence: 1 MSRYLRPNTSLFVRNVADD.....PKGKMKVLQYECTNITLV 216

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 1017041 seqs, 315518202 residues

Word size : 0

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

SPTREMBL 25:*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_virus:*
16: sp_bacterioplasmid:*
17: sp_archaeoplasmid:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	155	71.8	173	4 Q96G09	Q96G09 homo sapien
2	155	71.8	183	4 Q60572	Q60572 homo sapien
3	155	71.8	183	11 Q70307	Q70307 mus musculi
4	155	71.8	262	4 Q75494	Q75494 homo sapien
5	155	71.8	262	11 Q88468	Q88468 mus musculi
6	153	70.8	183	11 Q9ROT9	Q9ROT9 mus musculi
7	153	70.8	262	11 Q9R0U0	Q9R0U0 mus musculi
8	146	67.6	261	4 Q96P17	Q96P17 homo sapien
9	124	57.4	261	11 Q8CF51	Q8CF51 mus musculi
10	124	57.4	69	11 Q8CF51	Q8CF51 mus musculi
11	27	12.5	261	4 Q8WAF0	Q8WAF0 mus sapien
12	25	11.6	258	13 Q7ZUJ8	Q7ZUJ8 xenopus lae
13	21	9.7	261	4 Q8WAF0	Q8WAF0 mus sapien
14	19	8.8	248	13 Q7S2S0	Q7S2S0 brachydanio
15	12	5.6	773	5 Q9UJ79	Q9UJ79 drosophila
16	12	5.6	767	5 Q9VEJ1	Q9VEJ1 drosophila

17	9	4.2	226	10 Q8LQW8	Q8LQW8 oryza sativ
18	9	4.2	236	4 Q96JW1	Q96JW1 homo sapien
19	9	4.2	283	4 Q8TF00	Q8TF00 homo sapien
20	9	4.2	299	4 Q9Y4N3	Q9Y4N3 homo sapien
21	9	4.2	312	4 Q8NAV1	Q8NAV1 homo sapien
22	9	4.2	312	6 Q8HXH6	Q8HXH6 macaca fasc
23	9	4.2	312	11 Q8BYV2	Q8BYV2 mus musculi
24	9	4.2	347	10 Q9SU74	Q9SU74 arabidopsis
25	9	4.2	521	5 Q9VDR5	Q9VDR5 drosophila
26	9	4.2	523	13 Q803H1	Q803H1 brachydanio
27	9	4.2	805	4 Q8TF01	Q8TF01 homo sapien
28	9	4.2	872	5 Q8MZH3	Q8MZH3 drosophila
29	9	4.2	872	5 Q9VGV5	Q9VGV5 drosophila
30	8	3.7	48	5 Q86LB9	Q86LB9 entamoeba h
31	8	3.7	52	12 Q91GG6	Q91GG6 epiphyas po
32	8	3.7	54	12 Q8UR11	Q8UR11 phthorimaea
33	8	3.7	55	12 Q79U33	Q79U33 adoxophyes
34	8	3.7	56	12 Q9DVM6	Q9DVM6 pluteia xy
35	8	3.7	60	12 Q9PYU9	Q9PYU9 xestia c-ni
36	8	3.7	79	12 Q7T5F8	Q7T5F8 anthracis p
37	8	3.7	87	12 Q91EY8	Q91EY8 cydia pomon
38	8	3.7	118	5 Q8MSW0	Q8MSW0 caenorhabdi
39	8	3.7	129	11 Q9CVD6	Q9CVD6 mus musculi
40	8	3.7	141	6 Q8WMQ1	Q8WMQ1 macaca mula
41	8	3.7	146	6 Q8WMQ0	Q8WMQ0 macaca mula
42	8	3.7	153	5 Q8MSW1	Q8MSW1 caenorhabdi
43	8	3.7	155	4 Q8WVZ2	Q8WVZ2 homo sapien
44	8	3.7	155	11 Q8K194	Q8K194 mus musculi
45	8	3.7	160	11 Q9CM27	Q9CM27 mus musculi

ALIGNMENTS

RESULT 1

ID	Q96G09	PRELIMINARY;	PRT;	173 AA.
AC	Q96G09;			
DT	01-DEC-2001 (TREMURel. 19, Created)			
DT	01-DEC-2001 (TREMURel. 19, Last sequence update)			
DT	01-OCT-2003 (TREMURel. 25, Last annotation update)			
DE	TLS-associated serine-arginine protein 1.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.			
OX	NCBI_TaxID=9606;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=Uterus;			
RA	Strasbourg R.;			
RI	Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.			
DR	EMBL: BC010074; AAI10074.1; -			
DR	GO: GO:0003676; F:nucleic acid binding; IEA.			
DR	InterPro: IPR000504; RNA_rec_mot.			
DR	Pfam: PF00076; trm; 1.			
DR	SMART: SM00360; RRM; 1.			
DR	PROSITE: PSS0102; RRM; 1.			
DR	PROSITE: PSS0030; RRM_RNP_1; 1.			
SQ	SEQUENCE 173 AA; 21000 MW; 4AB40FBE24846495 CRC64;			

Query Match 71.8%; Score 155; DB 4; Length 173;
Best local Similarity 100.0%; Pred. No. 1.3e-158;
Matches 155; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1 MSRYLRPNTSLFVRNVADTRSEDIARRERGGPIVDVYPIIDFYTRRRRGPAVQFED 60	
DB	1 MSRYLRPNTSLFVRNVADTRSEDIARRERGGPIVDVYPIIDFYTRRRRGPAVQFED 60	
QY	61 VPAEDALHLIDKRWICGRQIEITQFAGDGTGNOMKAKGGRVYSSRRVDDRYRRSR 120	
DB	61 VPAEDALHLIDKRWICGRQIEITQFAGDGTGNOMKAKGGRVYSSRRVDDRYRRSR 120	
QY	121 SRSEYRRRSRRSFDYNYRRSYSPNRSRPTGRDR 155	

Db 121 SRSYERRSRSPDYNNRYSRPNRSRPTGRPR 155

RESULT 2

060572 PRELIMINARY; PRT; 183 AA.

AC 060572;

DC 01-AUG-1998 (TREMBLrel. 07, Created)

DT 01-AUG-1998 (TREMBLrel. 07, Last sequence update)

DT 01-JUN-2003 (TREMBLrel. 24, Last annotation update)

DE TLS-associated protein TASR (Hypothetical protein FLJ10424)

DE TLS-associated serine-arginine protein 1 (TLS-associated protein TASR-1) (TLS-associated SR protein 1) (Splicing factor SRp38-2).

GN TASR.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

OX NCBI_TaxID=9606;

RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE=Leukemia;

RA Yang L., Embree L., Tsai S., Hickestein D.D.;

RL Submitted (FEB-1998) to the EMBL/GenBank/DBJ databases.

RN [2]

RP SEQUENCE FROM N.A.

RA Isogai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y., Nishikawa T., Nagai K., Sugano S., Shiratori A., Sudo H., Wagatsuma M., Hoshitani T., Kaku Y., Kodaira H., Sugawara M., Takahashi M., Chiba Y., Ishida S., Murakawa K., Ono Y., Takiguchi S., Watanabe S., Kimura K., Murakami K., Ishii S., Kawai Y., Saito K., Yamamoto J., Wakamatsu K., Nakamura Y., Nagahara K., Masuho Y., Ninomiya K., Iwayanagi T.;

RA "NEDO human cDNA sequencing project.;"

RT Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.

RN [3]

RP SEQUENCE FROM N.A.

RC TISSUE=Placenta;

RA Strausberg R.;

RL Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.

RN [4]

RP SEQUENCE FROM N.A.

RA Clinton J.M., Chansky H.A., Odell D.D., Zielinska-Kwiatkowska A., Yang L.;

RA "Characterization and expression of the human gene encoding two TLS-associated serine-arginine (TASR) proteins.;"

RT Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.

RN [5]

RP SEQUENCE FROM N.A.

RX MEDLINE=98447613; PubMed=9774382;

RA Yang L., Embree L.J., Tsai S., Hickestein D.D.;

RT "Oncoprotein TLS interacts with serine-arginine proteins involved in RNA splicing.;"

RL J. Biol. Chem. 273:27761-27764(1998).

RN [6]

RP SEQUENCE FROM N.A.

RX MEDLINE=2042028; PubMed=10779324;

RA Yang L., Embree L.J., Hickestein D.D.;

RT "TUS-ERG leukemia fusion protein inhibits RNA splicing mediated by serine-arginine proteins.;"

RL Mol. Cell. Biol. 20:3345-3354(2000).

RN [7]

RP SEQUENCE FROM N.A.

RA Clinton J.M., Chansky H.A., Odell D.D., Zielinska-Kwiatkowska A., Yang L.;

RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.

RN [8]

RP SEQUENCE FROM N.A.

RX MEDLINE=2206915; PubMed=12419250;

RA Shin C., Manley J.L.;

RT "The SR Protein SRP38 Represses Splicing in M Phase Cells.;"

RL Cell 111:407-417(2002).

DR EMBL; AF047448; AAC70918.1; -.

DR EMBL; AK001286; BA391601.1; -

DR EMBL; BC001107; AA01107.1; -

DR EMBL; AF419331; AAL16665.1; -

DR EMBL; AY048592; AAL06098.1; -

DR EMBL; AY150181; AAN65381.1; -

DR HSSP; P11940; ICVJ.

DR GO; GO:0008248; F:pre-mRNA splicing factor activity; TAS.

DR GO; GO:0006397; P:mRNA processing; TAS.

DR GO; GO:0000245; P:spliceosome assembly; TAS.

DR InterPro; IPR000504; RNA_rec_mot.

DR Pfam; PF00076; rtm; 1.

DR SMART; SM00360; RRM; 1.

DR PROSITE; PS0102; RRM; 1.

DR PROSITE; PS00030; RRM_RNP_1; 1.

DR Hypothetical protein.

SQ SEQUENCE 183 AA; 22222 MW; 4AA87CA9B51A131 CRC64;

Query Match 71.8%; Score 155; DB 4; Length 183;

Best Local Similarity 100.0%; Pred. No. 1,4e-158;

Matches 155; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSRYLRPNTSLFVNNVADDTRESDLRRFGRYGVIVDYPVPLDFTYRPRGFAVYQFED 60

DB 1 MSRYLRPNTSLFVNNVADDTRESDLRRFGRYGVIVDYPVPLDFTYRPRGFAVYQFED 60

QY 61 VRDAEDALHNDLRKWCIGROIEIQPAQGRKTPNQMKAKGRNYSSGRYDPRYRGR 120

DB 61 VRDAEDALHNDLRKWCIGROIEIQPAQGRKTPNQMKAKGRNYSSGRYDPRYRGR 120

QY 121 SRSYERRSRSPDYNNRYSRPNRSRPTGRPR 155

DB 121 SRSYERRSRSPDYNNRYSRPNRSRPTGRPR 155

RESULT 3

070307 PRELIMINARY; PRT; 183 AA.

AC 070307;

DT 01-AUG-1998 (TREMBLrel. 07, Created)

DT 01-AUG-1998 (TREMBLrel. 07, Last sequence update)

DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)

DE TLS-associated protein with SR repeats.

GN NSSR.

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Scuriognathi; Muridae; Murinae; Mus.

OX NCBI_TaxID=10090;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=98447613; PubMed=9774382;

RA Yang L., Embree L.J., Tsai S., Hickestein D.D.;

RT "Oncoprotein TLS interacts with serine-arginine proteins involved in RNA splicing.;"

RL J. Biol. Chem. 273:27761-27764(1998).

DR EMBL; AF042383; AAC70916.1; -.

DR HSSP; P11940; ICVJ.

DR MGD; MGI:1333805; Nssr.

DR GO; GO:0003676; F:nucleic acid binding; IEA.

DR InterPro; IPR000504; RNA_rec_mot.

DR Pfam; PF00076; rtm; 1.

DR PROSITE; PS0102; RRM; 1.

DR PROSITE; PS00030; RRM_RNP_1; 1.

SQ SEQUENCE 183 AA; 22222 MW; 4AA87CA9B51A131 CRC64;

Query Match 71.8%; Score 155; DB 11; Length 183;

Best Local Similarity 100.0%; Pred. No. 1,4e-158;

Matches 155; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSRYLRPNTSLFVNNVADDTRESDLRRFGRYGVIVDYPVPLDFTYRPRGFAVYQFED 60

DB 1 MSRYLRPNTSLFVNNVADDTRESDLRRFGRYGVIVDYPVPLDFTYRPRGFAVYQFED 60

QY 61 VRDAEDALHNDLRKWCIGROIEIQPAQGRKTPNQMKAKGRNYSSGRYDPRYRGR 120

```

Db      61 VRDAEDALHNDLRKWMICGRQIEIQFAQGDRTPTQMKAKGGRNYSSSRDYDDYDRYRSR 120
Qy      121 SRSYERRRSRSPDYVNRYSRSPRNSRPTGRPR 155
      121 SRSYERRRSRSPDYVNRYSRSPRNSRPTGRPR 155

RESULT 4
ID 075494 PRELIMINARY; PRT: 262 AA.
AC 075494;
DT 01-NOV-1998 (TrEMBLrel. 08, Created)
DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE TLS-associated protein TASR-2 (TLS-associated serine-arginine protein
  2) (TLS-associated SR protein 2) (SRp40) (Splicing factor
  SRp38).
GN TASR.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Leukemia;
RA Yang L., Embree L., Hickstein D.D., Tsai S.;
RT "Molecular cloning of human TASR-2, a TLS-associated protein with Ser-
  Arg repeats.";
RL Submitted (MAY-1998) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Muscle;
RA Strausberg R.;
RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC MEDLINE=98447613; PubMed=9774382;
RA Yang L., Embree L.J., Tsai S., Hickstein D.D.;
RT "Oncoprotein TLS interacts with serine-arginine proteins involved in
  RNA splicing.";
RL J. Biol. Chem. 273:27761-27764(1998).
RN [4]
RP SEQUENCE FROM N.A.
RC MEDLINE=20242028; PubMed=10779324;
RA Yang L., Embree L.J., Hickstein D.D.;
RT "TLS-ERG leukemia fusion protein inhibits RNA splicing mediated by
  serine-arginine proteins.";
RL Mol. Cell. Biol. 20:3345-3354(2000).
RN [5]
RP SEQUENCE FROM N.A.
RA Clinton J.M., Chansky H.A., Odell D.D., Zielinska-Kwiatkowska A.,
  Yang L.;
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
RN [6]
RP SEQUENCE FROM N.A.
RC MEDLINE=21623667; PubMed=11684676;
RA Comper A.E., Caceres J.F., Mayeda A., Screaton G.R.;
RT "Serine-Arginine (SR) Protein-like Factors That Antagonize Authentic
  SR Proteins and Regulate Alternative Splicing.";
RL J. Biol. Chem. 276:48908-48914(2001).
RN [7]
RP SEQUENCE FROM N.A.
RC MEDLINE=22306915; PubMed=12419250;
RA Shin C., Manley J.L.;
RT "The SR Protein SRP38 Represses Splicing in M Phase Cells.";
RL Cell 111:407-417(2002).
DR EMBL: AF067730; AAC26727.1; -
DR EMBL: BC005039; AAH05039.1; -
DR EMBL: AY048592; AAH06099.1; -
DR EMBL: AF449427; AAL57514.1; -
DR EMBL: AY150180; AAN65380.1; -
DR HSSP: P11940; ILCVT.

```

```

DR Genew; HGNC:16713; FUSIP1.
DR GO; GO:0003676; F:nucleic acid binding; IEA.
DR InterPro; IPR000504; RNA_rec_mot.
DR Pfam; PF00076; rrm; 1.
DR SMART; SM00360; RRM; 1.
DR PROSITE; PSS0102; RRM; 1.
DR PROSITE; PS00030; RRM_RNP_1; 1.
SQ SEQUENCE 262 AA; 31300 MW; 205F95D36CBFB4 CRC64;

Query Match 71.8%; Score 155; DB 4; Length 262;
Best Local Similarity 100.0%; Pred. No. 1.9e-158;
Matches 155; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1 MSRYLRPNTSLFVRNVAADTRSEDLRRERXGPIVDVYVPLDFYTRRPRGFAYVQED 60
      1 MSRYLRPNTSLFVRNVAADTRSEDLRRERXGPIVDVYVPLDFYTRRPRGFAYVQED 60
Db      61 VRDAEDALHNDLRKWMICGRQIEIQFAQGDRTPTQMKAKGGRNYSSSRDYDDYDRYRSR 120
Qy      61 VRDAEDALHNDLRKWMICGRQIEIQFAQGDRTPTQMKAKGGRNYSSSRDYDDYDRYRSR 120
Db      61 VRDAEDALHNDLRKWMICGRQIEIQFAQGDRTPTQMKAKGGRNYSSSRDYDDYDRYRSR 120
Qy      121 SRSYERRRSRSPDYVNRYSRSPRNSRPTGRPR 155
      121 SRSYERRRSRSPDYVNRYSRSPRNSRPTGRPR 155
Db      121 SRSYERRRSRSPDYVNRYSRSPRNSRPTGRPR 155

RESULT 5
ID 088468 PRELIMINARY; PRT: 262 AA.
AC 088468;
DT 01-NOV-1998 (TrEMBLrel. 08, Created)
DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE TLS-associated protein TASR-2 (FUS-associated protein with
  serine-arginine repeats) (Similar to neural-salient
  DE serine/arginine-rich).
GN NSR OR FASR.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murine; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC MEDLINE=98447613; PubMed=9774382;
RA Yang L., Embree L., Tsai S., Hickstein D.D.;
RT "Molecular cloning of TASR-2, a TLS-associated protein with Ser-Arg
  repeats.";
RL Submitted (APR-1998) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Embryo;
RX MEDLINE=21085660; PubMed=11217951;
RA Kawai J., Shingawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
  Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
  Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamataka I.,
  Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
  Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
  Riechmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
  Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
  Schriml L.M., Staudt F., Suzuki R., Tomita M., Wagner J., Washio T.,
  Sakai K., Okido T., Furuno M., Aono H., Balderelli R., Barsh G.,
  Blake J., Boiffelli D., Bojunga N., Carlini P., de Bonaldo M.F.,
  Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
  Gustincich S., Hill D., Hofmann M., Hune D.A., Kamiya M., Lee N.H.,
  Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
  Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
  Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
  Suzuki H., Toyooka K., Wang K.H., Weitz C., Whitaker C., Wilming L.,
  Wyshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohatsu S.,
  Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
RN [3]
RP SEQUENCE FROM N.A.

```

```
RC STRAIN=C57BL/6; TISSUE=Brain;
RA Strausberg R.;
RL Submitted (JAN-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF060490; AAC26715.1; -.
DR EMBL; AK014345; BAB29286.1; -.
DR EMBL; BC043060; AAB43060.1; -.
DR HSSP; P11940; ICVU.
DR MGD; MGI:1333805; Nsr.
DR GO; GO:0003676; F:nucleic acid binding; IEA.
DR InterPro; IPR000504; RNA_rec_mot.
DR Pfam; PF00076; rrm; 1.
DR PROSITE; PS0102; RRM; 1.
DR PROSITE; PS00030; RRM_RNP_1; 1.
SQ SEQUENCE 262 AA; 3130 MW; 205F95D36CBFB4 CRC64;

Query Match 71.8%; Score 155; DB 11; Length 262;
Best Local Similarity 100.0%; Pred. No. 1.9e-158;
Matches 155; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSRYLRPNTSLFVNNVADDTRESEDLRRFGRGYPVIVVYVLDYTRPRGFAVVOED 60
DB 1 MSRYLRPNTSLFVNNVADDTRESEDLRRFGRGYPVIVVYVLDYTRPRGFAVVOED 60

QY 61 VRDAEDALHNDLRKWCGRQIEIQFAGCDRKTNPOMKAKEGRNVYSSSRYYDDYDRYRSR 120
DB 61 VRDAEDALHNDLRKWCGRQIEIQFAGCDRKTNPOMKAKEGRNVYSSSRYYDDYDRYRSR 120

QY 121 SRSYERRRSRSGSPDYNNRSYSPNRSRPTGR 155
DB 121 SRSYERRRSRSGSPDYNNRSYSPNRSRPTGR 155

RESULT 6
Q9ROT9 PRELIMINARY; PRT; 183 AA.
ID Q9ROT9
AC Q9ROT9;
RT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DE 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
DE Neutral specific sr protein NSSR 2.
GN NSSR OR NSSR 2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
CX NCBT_Taxid=10090;
RN 11
RP SEQUENCE FROM N.A.
RX MEDLINE=20051033; PubMed=10583508;
RA Komatsu M., Komimami E., Arahata K., Tsukahara T.;
RT "Cloning and characterization of two neutral-salient serine/arginine-
RT rich (NSSR) proteins involved in the regulation of alternative
RT splicing in neurones.";
RL Genes Cells 4:593-608 (1999).
DR EMBL; AB015895; BAA35093.1; -.
DR HSSP; P11940; ICVU.
DR MGD; MGI:1333805; Nsr.
DR GO; GO:0003676; F:nucleic acid binding; IEA.
DR InterPro; IPR000504; RNA_rec_mot.
DR Pfam; PF00076; rrm; 1.
DR SMART; SM00360; RRM; 1.
DR PROSITE; PS0102; RRM; 1.
DR PROSITE; PS00030; RRM_RNP_1; 1.
SQ SEQUENCE 183 AA; 22167 MW; 4AA87CA9B57C157 CRC64;

Query Match 70.8%; Score 153; DB 11; Length 183;
Best Local Similarity 100.0%; Pred. No. 2e-156;
Matches 153; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSRYLRPNTSLFVNNVADDTRESEDLRRFGRGYPVIVVYVLDYTRPRGFAVVOED 60
DB 1 MSRYLRPNTSLFVNNVADDTRESEDLRRFGRGYPVIVVYVLDYTRPRGFAVVOED 60

QY 61 VRDAEDALHNDLRKWCGRQIEIQFAGCDRKTNPOMKAKEGRNVYSSSRYYDDYDRYRSR 120
DB 61 VRDAEDALHNDLRKWCGRQIEIQFAGCDRKTNPOMKAKEGRNVYSSSRYYDDYDRYRSR 120

QY 121 SRSYERRRSRSGSPDYNNRSYSPNRSRPTGR 155
DB 121 SRSYERRRSRSGSPDYNNRSYSPNRSRPTGR 155
```

```
DB 61 VRDAEDALHNDLRKWCGRQIEIQFAGCDRKTNPOMKAKEGRNVYSSSRYYDDYDRYRSR 120

QY 121 SRSYERRRSRSGSPDYNNRSYSPNRSRPTGR 153
DB 121 SRSYERRRSRSGSPDYNNRSYSPNRSRPTGR 153

RESULT 7
Q9ROU0 PRELIMINARY; PRT; 262 AA.
ID Q9ROU0
AC Q9ROU0;
RT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DE 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
DE Neutral specific sr protein NSSR 1.
GN NSSR OR NSSR 1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
CX NCBT_Taxid=10090;
RN 11
RP SEQUENCE FROM N.A.
RX MEDLINE=20051033; PubMed=10583508;
RA Komatsu M., Komimami E., Arahata K., Tsukahara T.;
RT "Cloning and characterization of two neutral-salient serine/arginine-
RT rich (NSSR) proteins involved in the regulation of alternative
RT splicing in neurones.";
RL Genes Cells 4:593-608 (1999).
DR EMBL; AB015894; BAA35092.1; -.
DR HSSP; P11940; ICVU.
DR MGD; MGI:1333805; Nsr.
DR GO; GO:0003676; F:nucleic acid binding; IEA.
DR InterPro; IPR000504; RNA_rec_mot.
DR Pfam; PF00076; rrm; 1.
DR SMART; SM00360; RRM; 1.
DR PROSITE; PS0102; RRM; 1.
DR PROSITE; PS00030; RRM_RNP_1; 1.
SQ SEQUENCE 262 AA; 31245 MW; EB32279A4B5B8353 CRC64;

Query Match 70.8%; Score 153; DB 11; Length 262;
Best Local Similarity 100.0%; Pred. No. 2.8e-156;
Matches 153; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSRYLRPNTSLFVNNVADDTRESEDLRRFGRGYPVIVVYVLDYTRPRGFAVVOED 60
DB 1 MSRYLRPNTSLFVNNVADDTRESEDLRRFGRGYPVIVVYVLDYTRPRGFAVVOED 60

QY 61 VRDAEDALHNDLRKWCGRQIEIQFAGCDRKTNPOMKAKEGRNVYSSSRYYDDYDRYRSR 120
DB 61 VRDAEDALHNDLRKWCGRQIEIQFAGCDRKTNPOMKAKEGRNVYSSSRYYDDYDRYRSR 120

QY 121 SRSYERRRSRSGSPDYNNRSYSPNRSRPTGR 153
DB 121 SRSYERRRSRSGSPDYNNRSYSPNRSRPTGR 153

RESULT 8
Q96P17 PRELIMINARY; PRT; 261 AA.
ID Q96P17
AC Q96P17;
RT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DE 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
DE T1S-associated protein T1SR-2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
CX NCBT_Taxid=9606;
RN 11
RP SEQUENCE FROM N.A.
RA Clinton J.M., Chansky H.A., Odell D.D., Zielinska-Kwiatkowska A.,
Yang L.;
```

RT "Characterization and expression of the human gene encoding two TLS-
RT associated serine-arginine (TASR) proteins."
RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF413332; AALL666.1; "-
DR GO; GO:0003676; F:nucleic acid binding; IEA.
DR InterPro; IPR000504; RNA_rec_mot.
DR Pfam; PF00076; rrm; 1.
DR SMART; SM00360; RRM; 1.
DR PROSITE; PS50102; RRM; 1.
DR PROSITE; PS00030; RRM_RNP_1; 1.
SQ SEQUENCE 261 AA; 31213 MW; A07499B1D4C7570 CRC64;

Query Match 67.6%; Score 146; DB 4; Length 261;
Best Local Similarity 100.0%; Pred. No. 9.7e-149;
Matches 146; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSRYLRPNTSLFVNVAADTRSEDLRRFGRYGVIVVYVLDFTYTRRPGFAYVQED 60
DB 1 MSRYLRPNTSLFVNVAADTRSEDLRRFGRYGVIVVYVLDFTYTRRPGFAYVQED 60
QY 61 VRDAEDALHNDKRMICGRQIEIQAQGDRTKTPNQMKAKGRNYSRSSRYDDYRRRSR 120
DB 61 VRDAEDALHNDKRMICGRQIEIQAQGDRTKTPNQMKAKGRNYSRSSRYDDYRRRSR 120
QY 121 SRSYERRRSRSPDYNYRRSGYSPEN 146
DB 121 SRSYERRRSRSPDYNYRRSGYSPEN 146

RESULT 9
Q8CFZ1 PRELIMINARY; PRT; 261 AA.
ID Q8CFZ1
AC Q8CFZ1
DT 01-MAR-2003 (TREMBlrel. 23, Created)
DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE Similar to FUS interacting protein (Serine-arginine rich) 1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX NCBI_Taxid=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=FVB/N-3;
RA Strausberg R.;
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC037591; AAH37591.1; "-
DR GO; GO:0003676; F:nucleic acid binding; IEA.
DR InterPro; IPR000504; RNA_rec_mot.
DR Pfam; PF00076; rrm; 1.
DR SMART; SM00360; RRM; 1.
DR PROSITE; PS50102; RRM; 1.
DR PROSITE; PS00030; RRM_RNP_1; 1.
SQ SEQUENCE 261 AA; 31185 MW; C9500BF970FAAE1D CRC64;

Query Match 57.4%; Score 124; DB 11; Length 261;
Best Local Similarity 100.0%; Pred. No. 5.2e-125;
Matches 124; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 23 SEDLRREFGRYGVIVVYVLDFTYTRRPGFAYVQEDVDAEDALHNDKRMICGRQIE 82
DB 23 SEDLRREFGRYGVIVVYVLDFTYTRRPGFAYVQEDVDAEDALHNDKRMICGRQIE 82
QY 83 IQPAQGDRTKTPNQMKAKGRNYSRSSRYDDYRRRSRSGYSPDYNYRRSY 142
DB 83 IQPAQGDRTKTPNQMKAKGRNYSRSSRYDDYRRRSRSGYSPDYNYRRSY 142

QY 143 SPEN 146
DB 143 SPEN 146

RESULT 10

Q8CF51
ID Q8CF51 PRELIMINARY; PRT; 69 AA.
AC Q8CF51
DT 01-MAR-2003 (TREMBlrel. 23, Created)
DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE Neural-salient serine/arginine-rich.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_Taxid=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Cerebellum;
RX MEDLINE=22354683; PubMed=12466851;
RA The FANTOM Consortium,
RA the RIKEN Genome Exploration Research Group Phase I & II Team;
RT "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs."
RL Nature 420:563-573 (2002).
DR EMBL; AK005295; BAC25110.1; "-
DR GO; GO:0003676; F:nucleic acid binding; IEA.
DR InterPro; IPR000504; RNA_rec_mot.
DR Pfam; PF00076; rrm; 1.
DR SMART; SM00360; RRM; 1.
DR PROSITE; PS50102; RRM; 1.
DR PROSITE; PS00030; RRM_RNP_1; 1.
SQ SEQUENCE 69 AA; 8347 MW; AD2322054AE747B5 CRC64;

Query Match 26.4%; Score 57; DB 11; Length 69;
Best Local Similarity 100.0%; Pred. No. 2.8e-53;
Matches 57; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSRYLRPNTSLFVNVAADTRSEDLRRFGRYGVIVVYVLDFTYTRRPGFAYVQ 57
DB 1 MSRYLRPNTSLFVNVAADTRSEDLRRFGRYGVIVVYVLDFTYTRRPGFAYVQ 57

RESULT 11
Q8WYF0 PRELIMINARY; PRT; 261 AA.
ID Q8WYF0
AC Q8WYF0
DT 01-MAR-2002 (TREMBlrel. 20, Created)
DT 01-MAR-2002 (TREMBlrel. 20, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE SRP35.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
OX NCBI_Taxid=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21623667; PubMed=11684676;
RA Cowper A.E., Caceres J.F., Mayeda A., Sreaton G.R.;
RT "Serine-Arginine (SR) Protein-like Factors That Antagonize Authentic
RT SR Proteins and Regulate Alternative Splicing."
RL J. Biol. Chem. 276:48908-48914 (2001).
DR EMBL; AF449428; AAL57515.1; "-
DR GO; GO:0003676; F:nucleic acid binding; IEA.
DR InterPro; IPR000504; RNA_rec_mot.
DR Pfam; PF00076; rrm; 1.
DR SMART; SM00360; RRM; 1.
DR PROSITE; PS50102; RRM; 1.
DR PROSITE; PS00030; RRM_RNP_1; 1.
SQ SEQUENCE 261 AA; 30512 MW; C1473A5760462B54 CRC64;

Query Match 12.5%; Score 27; DB 4; Length 261;
Best Local Similarity 100.0%; Pred. No. 2.1e-20;
Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 42 PLDFYTRRPRGFGYVQEDVRDAEDAL 68
DB 42 PLDFYTRRPRGFGYVQEDVRDAEDAL 68

```

RESULT 12
07ZTU8      PRELIMINARY;      PRT;      258 AA.
AC 07ZTU8:
DT 01-JUN-2003 (TRENBLrel. 24, Created)
DT 01-JUN-2003 (TRENBLrel. 25, Last sequence update)
DT 01-OCT-2003 (TRENBLrel. 25, Last annotation update)
DE Similar to FUS interacting protein (Serine-arginine rich) 1.
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipiloidea; Pipidae;
OC Xenopodinae; Xenopus.
OX NCBI_TaxID=8335;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Embryo;
RA Klein S., Strausberg R.;
RL Submitted (FEB-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC046695; AAH46695.1;
DR GO; GO:0003676; F:mucleic acid binding; IEA.
DR InterPro; IPR000504; RNA_rec_mot.
DR Pfam; PF00076; rrm; 1.
DR SMART; SM00360; RRM; 1.
DR PROSITE; PS50102; RRM; 1.
DR PROSITE; PS00030; RRM_RNP_1; 1.
SQ SEQUENCE 258 AA; 30285 MW; 8436A11BE3630BAB CRC64;

Query Match      11.6%; Score 25; DB 13; Length 258;
Best Local Similarity 100.0%; Pred. No. 3e-18;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 48 RPRGFAYVQFEDVDAEDALHND 72
DB 48 RPRGFAYVQFEDVDAEDALHND 72

RESULT 13
08MW25      PRELIMINARY;      PRT;      261 AA.
AC 08MW25:
DT 01-MAR-2002 (TRENBLrel. 20, Created)
DT 01-MAR-2002 (TRENBLrel. 20, Last sequence update)
DT 01-OCT-2003 (TRENBLrel. 25, Last annotation update)
DE Similar to FUS interacting protein (Serine-arginine rich) 2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiinae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Testis;
RA Strausberg R.;
RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC021715; AAH21715.1;
DR GO; GO:0003676; F:mucleic acid binding; IEA.
DR InterPro; IPR000504; RNA_rec_mot.
DR Pfam; PF00076; rrm; 1.
DR SMART; SM00360; RRM; 1.
DR PROSITE; PS50102; RRM; 1.
DR PROSITE; PS00030; RRM_RNP_1; 1.
SQ SEQUENCE 261 AA; 30454 MW; 2420CBDE4284755 CRC64;

Query Match      9.7%; Score 21; DB 4; Length 261;
Best Local Similarity 100.0%; Pred. No. 6.3e-14;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 42 PLDFYTRRRGFAVYQFEDVR 62
DB 42 PLDFYTRRRGFAVYQFEDVR 62

```

```

RESULT 14
07SZSO      PRELIMINARY;      PRT;      248 AA.
AC 07SZSO:
DT 01-OCT-2003 (TRENBLrel. 25, Created)
DT 01-OCT-2003 (TRENBLrel. 25, Last sequence update)
DT 01-OCT-2003 (TRENBLrel. 25, Last annotation update)
DE Hypothetical protein.
OS Brachydanio rerio (Zebrafish) (Danio rerio).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
OC Cyprinidae; Danio.
OX NCBI_TaxID=7955;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Embryo;
RX MEDLINE=22388257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diachenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Uscdin T.B., Toshiyuki S., Carninci P., Frange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullany S.J.,
RA Bosak S.A., McKernan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,
RA Whitting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butlerfield Y.S.,
RA Krzyzanski M.I., Skalska U., Smailus D.E., Schnerch A., Schein J.E.,
RA Jones S.J., Maira M.A.;
RT "Genes and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences."
RT Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Embryo;
RA Strausberg R.;
RL Submitted (AUG-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC056275; AAH56275.1;
KW Hypothetical protein.
SQ SEQUENCE 248 AA; 29357 MW; C6E243AC3D459335 CRC64;

Query Match      8.8%; Score 19; DB 13; Length 248;
Best Local Similarity 100.0%; Pred. No. 8.6e-12;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 57 QFEDVDAEDALHNDKRW 75
DB 57 QFEDVDAEDALHNDKRW 75

RESULT 15
09UJ9       PRELIMINARY;      PRT;      773 AA.
AC 09UJ9:
DT 01-MAY-2000 (TRENBLrel. 13, Created)
DT 01-MAY-2000 (TRENBLrel. 13, Last sequence update)
DT 01-OCT-2003 (TRENBLrel. 25, Last annotation update)
DE SFI protein.
GN SFI OR CG5836.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RA Wazirout R., Puoti A., Kramer A.;
RT "Splicing factor SFI from Drosophila and Caenorhabditis contains an N-

```

RT terminal domain not present in the mammalian or yeast homologues."
RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: CONTAINS 1 KH DOMAIN.
DR EMBL; AJ243904; CAB64937.1; -.
DR FlyBase; FBgn0025571; SPL.
DR GO; GO:0003676; F:nucleic acid binding; IEA.
DR InterPro; IPR004087; KH_dom.
DR InterPro; IPR004088; KH_type_1.
DR InterPro; IPR01878; Znf_CCHC.
DR Pfam; PF00013; KH; 1.
DR Pfam; PF00098; zf-CCHC; 1.
DR PRINTS; PR00939; C2HCZNFINGER.
DR SMART; SM00322; KH; 1.
DR SMART; SM00343; Znf_C2HC; 2.
DR PROSITE; PS50084; KH_TYPR_1; 1.
DR PROSITE; PS50158; ZF_CCHC; 1.
SQ SEQUENCE 773 AA; 86526 MW; BFFCD08AE4755098 CRC64;

Query Match 5.6%; Score 12; DB 5; Length 773;
Best local Similarity 100.0%; Pred.No. 0.00085;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 117 RRSRSRSYERR 128
DB 145 RRSRSRSYERR 156

Search completed: July 22, 2004, 17:20:03
Job time : 41 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 29, 2004, 03:15:17 ; Search time 759 Seconds

(without alignments)
10477.769 Million cell updates/sec

Title: US-09-787-491B-18

Sequence: 1 gccctctgcgcgtgtgtgtc.....tatcaaaaaaaaaaaaaa 1872

Scoring table:

Gapop 60.0 , Gapext 60.0

Searched: 3373863 seqs, 2124099041 residues

Word size : 0

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database :

N Geneseg 29Jan04:*
1: genesegn1980s:*
2: genesegn1990s:*
3: genesegn2000s:*
4: genesegn2001as:*
5: genesegn2001bs:*
6: genesegn2002as:*
7: genesegn2003as:*
8: genesegn2003bs:*
9: genesegn2003cs:*
10: genesegn2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1872	100.0	1872	3	AAAI2401
2	1872	100.0	1872	3	AAZ51250
3	1716	91.7	1838	6	ABL87934
4	1483	79.2	1836	4	AAH14124
5	940	50.2	1326	3	AAZ51256
6	786	42.0	2473	7	ADA52465
7	709	37.9	1382	5	ABV25632
8	674	36.0	2247	4	AAH14349
9	509	27.2	702	4	AAH06844
10	509	27.2	2918	4	AAH14554
11	460	24.6	550	6	ABL87933
12	435	23.2	716	5	AAH79415
13	435	23.2	716	5	AAH79415
14	430	23.0	737	5	AAH79415
15	426	22.4	2407	5	ABV24425
16	420	22.4	881	4	AAH05638
17	394	21.0	882	4	AAH06231
18	370	19.8	392	5	AAH06231
19	306	16.3	392	5	AAH06231
20	306	16.3	392	5	AAH06231
21	266	13.2	504	9	ADD49742
22	247	13.2	247	9	AAH06231
23	228	12.2	351	2	AAH06231

C	24	187	10.0	264	4	AAH19930
C	25	149	8.0	541	5	ABV38666
C	26	130	6.9	301	5	ABV08770
C	27	100	5.3	299	4	AAK53699
C	28	93	5.0	442	4	AAH10350
C	29	63	3.4	741	4	AAH196799
C	30	50	2.7	50	6	ABZ07375
C	31	50	2.7	412	7	ABH35695
C	32	29	1.5	907	4	AAH04422
C	33	29	1.5	944	7	ABH34763
C	34	29	1.5	955	4	ABH58636
C	35	29	1.5	955	8	ABH48614
C	36	29	1.5	989	7	AAH54748
C	37	26	1.4	443	9	ABH51833
C	38	23	1.2	402	4	AAH20448
C	39	23	1.2	2483	9	ABH37612
C	40	22	1.2	285	5	ABH61170
C	41	22	1.2	332	6	ABH63404
C	42	22	1.2	422	6	ABH45001
C	43	22	1.2	470	9	ADD49792
C	44	22	1.2	478	9	ABH51222
C	45	22	1.2	486	6	ABH44763

ALIGNMENTS

RESULT 1	AAAI2401	standard; cDNA; 1872 BP.
XX	AAAI2401;	
AC	AAAI2401;	
XX	25-JUL-2000	(first entry)
DE	cdna encoding a human RNA-associated protein.	
XX		
KW	Human; RNA-associated protein; cell proliferation; cancer; inflammation;	
KW	immune response; reproductive disorder; acinic keratosis;	
KW	atherosclerosis; arteriosclerosis; bursitis; cirrhosis; hepatitis;	
KW	mixed connective tissue disease; myelofibrosis; primary thrombocythemia;	
KW	paroxysmal nocturnal hemoglobinuria; polycythemia vera; psoriasis;	
XX	trauma; ss.	
OS	Homo sapiens.	
XX		
FH	Key	Location/Qualifiers
FT	CDS	94..744
FT		/*tag= a
FT		/product= "RNA-associated protein"
XX		
PN	WC000015799-A2.	
XX		
PD	23-MAR-2000.	
XX		
PF	17-SEP-1999;	99MO-US021688.
XX		
PR	17-SEP-1998;	98US-00156039.
PR	22-SEP-1998;	98US-00158720.
PR	04-NOV-1998;	98US-00186815.
PR	08-APR-1999;	99US-0128660P.
XX		
PA	(INCY-) INCYTE PHARM INC.	
XX		
PI	Tang YF, Corley NC, Guegler KJ, Gorgone GA, Patterson C;	
PI	Hillman JL, Baughn WR, Lal P, Azimzai Y, Yue H, Yang J;	
XX		
DR	WPI: 2000-271437/23.	
XX		
DR	P-PSDB; AAH4429.	
XX		
PT	New polypeptides and polynucleotides, useful for preventing and treating	
PT	a disorder associated with increased or decreased expression of RNA	
PT	associated proteins.	

XX Claim 9; Page 112; 131bp; English.
PS
CC The present sequence encodes a human RNA-associated protein. The
CC expression of RNA-associated proteins is closely associated with
CC reproductive tissues, nervous tissues, cell proliferation including
CC cancer, inflammation and immune responses, and so they may be used for
CC diagnosis, treatment or prevention of cell proliferative,
CC immune/inflammatory disorders, and reproductive disorders. Diseases and
CC disorders which may be treated include actinic keratosis,
CC atherosclerosis, arteriosclerosis, buritis, cirrhosis, hepatitis, mixed
CC connective tissue disease, myelofibrosis, paroxysmal nocturnal
CC hemoglobinuria, polyarthralgia vera, psoriasis, primary thrombocythemia
CC and cancers, and trauma
XX
Sequence 1872 BP; 579 A; 334 C; 392 G; 567 T; 0 U; 0 Other;
Query Match 100.0%; Score 1872; DB 3; Length 1872;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1872; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GCCCTCTGCTGTGTGTCTGAGGCTCGGCGCTGAGCGCGGACGGTTGCTGAGCC 60
DB 1 GCCCTTACGCTGTGTGTCTGAGGCTCGGCGCTGAGCGCGGACGGTTGCTGAGCC 60
QY 61 CGTTAGTCCGCCCGCGGAGACACGCCCGCCATGTCCTGCTACCTGCTCCCGAC 120
DB 61 CGTTAGTCCGCCCGCGGAGACACGCCCGCCATGTCCTGCTACCTGCTCCCGAC 120
QY 121 ACGCTCTGCTGT 180
DB 121 ACGCTCTGCTGT 180
QY 181 TTGGTCTGT 240
DB 181 TTGGTCTGT 240
QY 241 CCAGAGAGATTTGCTTATGTTCAATTTGAGATGTTCTGATGCTGAGAGAGCTTTAAT 300
DB 241 CCAGAGAGATTTGCTTATGTTCAATTTGAGATGTTCTGATGCTGAGAGAGCTTTAAT 300
QY 301 AATTGGACAGAAAGTGATTTGTGAGCGGAGATTGAAATACAGTTGCCCGAGGGGAT 360
DB 301 AATTGGACAGAAAGTGATTTGTGAGCGGAGATTGAAATACAGTTGCCCGAGGGGAT 360
QY 361 CGAAAGACACCAATCAGATGAAAGCCAGAGAGAGAGATGTGTACGTTCTTCAAGC 420
DB 361 CGAAAGACACCAATCAGATGAAAGCCAGAGAGAGAGATGTGTACGTTCTTCAAGC 420
QY 421 TATGATGATTATGACAGATACAGAGCTTTAGAGCCGAAAGTTATGAAAGAGAGATCA 480
DB 421 TATGATGATTATGACAGATACAGAGCTTTAGAGCCGAAAGTTATGAAAGAGATCA 480
QY 481 AAGAAGCGCTTTGATTAACAATAGAGATGATAGTCTGATGCTGAGAAACAGTAGACCG 540
DB 481 AAGAAGCGCTTTGATTAACAATAGAGATGATAGTCTGATGCTGAGAAACAGTAGACCG 540
QY 541 ACTGGAAGACACAGCGGTAGAGAGCCATTCGACATGATATAGACCAATCTGAGCTGGA 600
DB 541 ACTGGAAGACACAGCGGTAGAGAGCCATTCGACATGATATAGACCAATCTGAGCTGGA 600
QY 601 ATACCCAGATACAGATTCGCTTACATTCAGAAAGAGTCTGAAAGCGGAAAGAAAGACC 660
DB 601 ATACCCAGATACAGATTCGCTTACATTCAGAAAGAGTCTGAAAGCGGAAAGAAAGACC 660
QY 661 AAAGAGAGGAGCTTCAAGAGACCAAGAGGCTGTGAGAGGCTGTGAGATGATCTGT 720
DB 661 AAAGAGAGGAGCTTCAAGAGACCAAGAGGCTGTGAGAGGCTGTGAGATGATCTGT 720
QY 721 ACGAATATTTTGAAGCTGTGCTGAGAAAGATTAAGAAATGTTATCGAAATACATGGAAT 780
DB 721 ACGAATATTTTGAAGCTGTGCTGAGAAAGATTAAGAAATGTTATCGAAATACATGGAAT 780

QY 781 AATTGAAGTCCCTTCAAGTTTGAAGAGTAAAGTAAATTTAGACAAATTAAGAAATTCAC 840
DB 781 AATTGAAGTCCCTTCAAGTTTGAAGAGTAAAGTAAATTTAGACAAATTAAGAAATTCAC 840
QY 841 TTGTGATCTTGAGAAATTAATCCCTAAATATGAATAGATTTATATGATATGATGGGTAC 900
DB 841 TTGTGATCTTGAGAAATTAATCCCTAAATATGAATAGATTTATATGATATGATGGGTAC 900
QY 901 AGGTCCATTAATTAATTTTGAAGAACTGAGAGTGTGAATATCAAGAGAGACGCAATAGT 960
DB 901 AGGTCCATTAATTAATTTTGAAGAACTGAGAGTGTGAATATCAAGAGAGACGCAATAGT 960
QY 961 CTCTTACAGTGCCTCTGT 1020
DB 961 CTCTTACAGTGCCTCTGT 1020
QY 1021 AATATTAAGTCCATTTTGTCCATTTATGAGCAATCTTGCTTTGTTATTTTGTGTGCC 1080
DB 1021 AATATTAAGTCCATTTTGTCCATTTATGAGCAATCTTGCTTTGTTATTTTGTGTGCC 1080
QY 1081 AGTGTCTTGTCTTAATATGATTTGCTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1140
DB 1081 AGTGTCTTGTCTTAATATGATTTGCTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1140
QY 1141 ATGCAATTTACATCTGTCTTAATCTCTTCCAGATTAATTCATTAATTTGACAT 1200
DB 1141 ATGCAATTTACATCTGTCTTAATCTCTTCCAGATTAATTCATTAATTTGACAT 1200
QY 1201 CCAGCTAAGAGGCGCCATCTCTTCTCACTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT 1260
DB 1201 CCAGCTAAGAGGCGCCATCTCTTCTCACTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT 1260
QY 1261 TATGAGCCCTTACCTGAGGAGCAATCATGTACTGATTAATTTGAGAAATATGATTAATTC 1320
DB 1261 TATGAGCCCTTACCTGAGGAGCAATCATGTACTGATTAATTTGAGAAATATGATTAATTC 1320
QY 1321 CCTTATTCAGTAATATGTCCTTACCTGAGCAATCTGATGATCAATCAATGATGCTCAT 1380
DB 1321 CCTTATTCAGTAATATGTCCTTACCTGAGCAATCTGATGATCAATCAATGATGCTCAT 1380
QY 1381 GTTTTGTGAGGTGTGTATTCATTAATTTTACCACTTGTGTGTGTGTGTGTGTGTGTGT 1440
DB 1381 GTTTTGTGAGGTGTGTATTCATTAATTTTACCACTTGTGTGTGTGTGTGTGTGTGTGT 1440
QY 1441 TAGAACAATATATAGATCAAGATTAATTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1500
DB 1441 TAGAACAATATATAGATCAAGATTAATTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1500
QY 1501 CCAGTATTTGATCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1560
DB 1501 CCAGTATTTGATCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1560
QY 1561 TTTATTTGTGAGTGTGTATTCATTAATTTTACCACTTGTGTGTGTGTGTGTGTGTGTGT 1620
DB 1561 TTTATTTGTGAGTGTGTATTCATTAATTTTACCACTTGTGTGTGTGTGTGTGTGTGTGT 1620
QY 1621 AAACACAGTCTGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1680
DB 1621 AAACACAGTCTGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1680
QY 1681 AGACTGGGTCTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1740
DB 1681 AGACTGGGTCTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1740
QY 1741 TTAGTATGTTCTGTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1800
DB 1741 TTAGTATGTTCTGTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1800
QY 1801 TACATTTTATAGGAGACATGTTCTGTGTATAGCAATTAATTAATTAATTAATTAATTA 1860
DB 1801 TACATTTTATAGGAGACATGTTCTGTGTATAGCAATTAATTAATTAATTAATTAATTA 1860
QY 1861 AAAAAAAAAA 1872

DB 1861 AAAAAAAAAA 1872

RESULT 2
AA251250
ID AA251250 standard; cDNA, 1872 BP.
XX
AC AA251250;
XX
DT 06-JUN-2000 (first entry)
XX
DE Human RNA-associated protein-1 (RNAP-1) encoding cDNA.
XX
KW RNA-associated protein; RNAP; human; clone 399781; cytosolic;
KW immunosuppressive; anti-inflammatory; keratolytic; neuroprotective;
KW antiarteriosclerotic; hepatotropic; antipsoriatic; virucide; anti-HIV;
KW antiallergic; antihemetic; antiarthritic; ophthalmological; autoimmune;
KW antimicrobial; cell proliferative disorder; inflammatory; cirrhosis;
KW actinic keratosis; bursitis; arteriosclerosis; artherosclerosis;
KW hepatitis; myelofibrosis; primary thrombocythemia; psoriasis; cancer;
KW mixed connective tissue disease; MCTD; HIV; uveitis; Crohn's disease;
KW allergy; rheumatoid arthritis; parasitic infection; ss.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FH CDS 94..744
FT /*tag= a
FT /product= "Human RNA-associated protein-1"
FT /note= "Derived from P170N0702 library"
XX
PN W020001171-A2.
XX
PD 02-MAR-2000.
XX
PE 20-AUG-1999; 99WO-US019361.
XX
PR 21-AUG-1998; 98US-0097550P.
PR 12-JAN-1999; 99US-0115639P.
XX
PA (INCY-) INCYTE PHARM INC.
XX
PI Hillman JL, Yue H, Tang YT, Corley NC, Guegler KJ, Gorgone GA;
PI Patterson C, Baughm MR, Lal P, Bandman O, Reddy R, Azimzai Y;
PI Shih JL, Yang J, Lu DM;
XX
DR WPI: 2000-237651/20.
DR P-PSDB: AAY70220.
XX
PT Human RNA-associated proteins useful in diagnosing, treating and
PT preventing cell proliferative, autoimmune, inflammatory and infectious
PT disorders.
XX
PS Claim 9; Page 105-106; 123pp; English.
XX
CC The present sequence is the cDNA encoding human RNA-associated protein-1
CC (RNAP-1), identified in Incyte clone 399781, derived from P170N0702
CC library. It is expressed in reproductive, nervous and cardiovascular
CC tissues. It has cytosolic, immunosuppressive, keratolytic,
CC antiinflammatory, antiarteriosclerotic, hepatotropic, virucide,
CC neuroprotective, antipsoriatic, anti-HIV, antiallergic, antihemetic,
CC antiarthritic, ophthalmological and antimicrobial activity. RNAP
CC antibodies are useful for diagnosis of diseases associated with altered
CC expression or activity of RNAP. It is used to treat cell proliferative,
CC autoimmune, inflammatory and infectious disorders, like actinic
CC keratosis, bursitis, arteriosclerosis, artherosclerosis, cirrhosis,
CC hepatitis, myelofibrosis, mixed connective tissue disease (MCTD),
CC psoriasis, primary thrombocythemia and cancer, HIV, allergies, rheumatoid
CC arthritis, uveitis, Crohn's disease, and bacterial, viral and parasitic
CC infections
XX
SQ Sequence 1872 BP; 579 A; 334 C; 392 G; 567 T; 0 U; 0 Other;

Query Match 100.0%; Score 1872; DB 3; Length 1872;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1872; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCCCTTAGCTGTGTGTGTGTGAGCTCGGCCCTGAGCGCGAGCTTGTGTGAGCC 60
DB 1 GCCCTTAGCTGTGTGTGTGTGAGCTCGGCCCTGAGCGCGAGCTTGTGTGAGCC 60
QY 61 CGTTAGTGCCTGCGCCGCGAGCAGCGCCGCGCATTCCGCTACCTGCTCCCAAC 120
DB 61 CGTTAGTGCCTGCGCCGCGAGCAGCGCCGCGCATTCCGCTACCTGCTCCCAAC 120
QY 121 ACGTCTGCTTGTGTCAGGAAAGTGGCGGAGCAGCGCTGAGAGCTTGGCGCTGAA 180
DB 121 ACGTCTGCTTGTGTCAGGAAAGTGGCGGAGCAGCGCTGAGAGCTTGGCGCTGAA 180
QY 181 TTGTGCTTATGCTCTTATGATGATGATGATGATGATGATGATGATGATGATGAT 240
DB 181 TTGTGCTTATGCTCTTATGATGATGATGATGATGATGATGATGATGATGATGAT 240
QY 241 CCAAGAGGATTTGCTTATGTTCAATTGAGGATTTGCTGATGCTGAAAGCGTTTACAT 300
DB 241 CCAAGAGGATTTGCTTATGTTCAATTGAGGATTTGCTGATGCTGAAAGCGTTTACAT 300
QY 301 AATTGACAGAAAGTGGATTTGTGACGGGAGATTGAATATACGTTTGGCCAGGGGAT 360
DB 301 AATTGACAGAAAGTGGATTTGTGACGGGAGATTGAATATACGTTTGGCCAGGGGAT 360
QY 361 CGAAAGACACCAATATGATGAAAGCCAGAGAGAGAGATGTATCAGTTCTTACGCG 420
DB 361 CGAAAGACACCAATATGATGAAAGCCAGAGAGAGAGATGTATCAGTTCTTACGCG 420
QY 421 TATGATGATTTATGACAGATTCAGACGTTCTTGAAGCCGAAAGTTTGAAGAGATCA 480
DB 421 TATGATGATTTATGACAGATTCAGACGTTCTTGAAGCCGAAAGTTTGAAGAGATCA 480
QY 481 AGAAGTGGTCTTTTGTATCAACTATGAAAGATCGTATGCTCTTAAAGACATGACCG 540
DB 481 AGAAGTGGTCTTTTGTATTCACACTATGAAAGATCGTATGCTCTTAAAGACATGACCG 540
QY 541 ACTGAAAGACACCGGCTAGAGAAAGCCATTCGACATATGATGACCAAACTGACGTGGA 600
DB 541 ACTGAAAGACACCGGCTAGAGAAAGCCATTCGACATATGATGACCAAACTGACGTGGA 600
QY 601 ATACCAAGTACAGTTCGCTTACTTACTTACCTTCAAGAAAGATCTGAAAGCGAAAGAAC 660
DB 601 ATACCAAGTACAGTTCGCTTACTTACTTACCTTCAAGAAAGATCTGAAAGCGAAAGAAC 660
QY 661 AAAAGAGGCGAGTTCAAGCAGCAAGAGGCGGTGGAGAGGCTGAGTATGATGATCTGT 720
DB 661 AAAAGAGGCGAGTTCAAGCAGCAAGAGGCGGTGGAGAGGCTGAGTATGATGATCTGT 720
QY 721 ACGAATATTTTGACTGCTGCTGCTGAAAAGATTAAGATGTTATCGAAAATACATGGAAT 780
DB 721 ACGAATATTTTGACTGCTGCTGCTGAAAAGATTAAGATGTTATCGAAAATACATGGAAT 780
QY 781 AATTGAAGTCCCTTCAAGTTTGAAGTGAACATTTTGAAGCAAAATTAAGAAATTTAC 840
DB 781 AATTGAAGTCCCTTCAAGTTTGAAGTGAACATTTTGAAGCAAAATTAAGAAATTTAC 840
QY 841 TTGTGCTTGTGGAAGCTATCCCTTAAATGTAATGATGATGATGATGATGATGATGAT 900
DB 841 TTGTGCTTGTGGAAGCTATCCCTTAAATGTAATGATGATGATGATGATGATGATGAT 900
QY 901 AGGTCAATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 960
DB 901 AGGTCAATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 960
QY 961 CTCTTACAGTCTCTGTTGT 1020
DB 961 CTCTTACAGTCTCTGTTGT 1020

```

QY 1021 AATAAAGTTCATTTTCCATTTATGGCAATCTTGCCTTTGTTATTTTGGTCC 1080
DB 1021 AATAAAGTTCATTTTCCATTTATGGCAATCTTGCCTTTGTTATTTTGGTCC 1080
QY 1081 AGTGTTCCTGCTTAATCATTTGCTTTGTCGACATCTGTTTATTTACTGTGACACAC 1140
DB 1081 AGTGTTCCTGCTTAATCATTTGCTTTGTCGACATCTGTTTATTTACTGTGACACAC 1140
QY 1141 ATGCAATTTACATCTGCTTAATCTCTTCCAGTAAATTCATTAATTTTGGACAT 1200
DB 1141 ATGCAATTTACATCTGCTTAATCTCTTCCAGTAAATTCATTAATTTTGGACAT 1200
QY 1201 CCAGTAAAGAGGCGCCATCTCTCCACCTCTTCCAGTAAATTCATTAATTTTGGACAT 1260
DB 1201 CCAGTAAAGAGGCGCCATCTCTCCACCTCTTCCAGTAAATTCATTAATTTTGGACAT 1260
QY 1261 TATTGAGCCCTTACTGCTGCGCAATCATTTGTAAGTAAATTTGAGAAATAGATATTC 1320
DB 1261 TATTGAGCCCTTACTGCTGCGCAATCATTTGTAAGTAAATTTGAGAAATAGATATTC 1320
QY 1321 CCTTATTCAGTAATGCTCTACTGAGCAATCTAGTATCATTTACAGTATGCGCTCAT 1380
DB 1321 CCTTATTCAGTAATGCTCTACTGAGCAATCTAGTATCATTTACAGTATGCGCTCAT 1380
QY 1381 GTTTTGTGAGGTGCTTATTCATTAATTTTACCAATTCGTATCATTTAATTA 1440
DB 1381 GTTTTGTGAGGTGCTTATTCATTAATTTTACCAATTCGTATCATTTAATTA 1440
QY 1441 TAGAACAATATATGATCAAGATAGATTAATTTGCTGATTCGCAATTTAAAGTAT 1500
DB 1441 TAGAACAATATATGATCAAGATAGATTAATTTGCTGATTCGCAATTTAAAGTAT 1500
QY 1501 CCAGTATTTGATCATTTATTAATTAATGAAAAATGATTTATCTGTATTAACCTGG 1560
DB 1501 CCAGTATTTGATCATTTATTAATTAATGAAAAATGATTTATCTGTATTAACCTGG 1560
QY 1561 TTATTTGTCAGTGCCTGTAATATCTAGAGTAAATTAATTTGCTGCTGCTCAC 1620
DB 1561 TTATTTGTCAGTGCCTGTAATATCTAGAGTAAATTAATTTGCTGCTGCTCAC 1620
QY 1621 AAACACATGCTAGATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1680
DB 1621 AAACACATGCTAGATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1680
QY 1681 AGACTGGGTCTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1740
DB 1681 AGACTGGGTCTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1740
QY 1741 TTAGTATGCTCTGTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1800
DB 1741 TTAGTATGCTCTGTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1800
QY 1801 TACATTTTATAGGGGACATGTTCTGTATTAAGCGAATTAATTAATTAATTAATTA 1860
DB 1801 TACATTTTATAGGGGACATGTTCTGTATTAAGCGAATTAATTAATTAATTAATTA 1860
QY 1861 AAAAAAAAAA 1872
DB 1861 AAAAAAAAAA 1872

```

RESULT 3

ABL87934 standard, DNA; 1838 BP.

ABL87934;

17-MAY-2002 (first entry)

Human ovarian cancer related DNA clone SEQ ID NO:10912.

Human; ovarian cancer; ovarian tumour; cytostatic; gene; ds.

XX

```

OS Homo sapiens.
XX
XX WO200192581-A2.
XX
XX 06-DEC-2001.
XX
XX 29-MAY-2001; 2001WO-US017756.
XX
XX 26-MAY-2000; 2000US-0207484P.
XX
XX (CORI-) CORIXA CORP.
XX
XX Algate PA, Harlocker SL, Jones R;
XX
XX WPI; 2002-122075/16.
XX
XX Composition for therapy and diagnosis of ovarian cancer comprising
PT polypeptide of a ovarian tumor polypeptide, polynucleotide encoding
PT polypeptide, antibody specific to polypeptide or T cell expressing
PT polypeptide.
XX
XX Claim 1; SEQ ID NO 10912; 489bp; English.
XX
XX The present invention describes a composition (I) comprising: carriers
CC and immunostimulants; and a polypeptide (II) of a ovarian tumour
CC polypeptide encoded by a polynucleotide (III) having a cDNA sequence (S1)
CC from the 10912 nucleotide sequences as given in ABL87934 to ABL87934,
CC (III) encoding (II) having a sequence (S2), a T cell population of (II),
CC or antigen presenting cells that express (II). (I) has cytostatic
CC activity. An oligonucleotide (IV) that hybridises to (S1) can be used for
CC detecting ovarian cancer in a patient's biological sample preferably
CC serum or ovarian tissue. The method comprises contacting a biological
CC sample from a patient with (IV), detecting the amount of polynucleotide
CC hybridising to (IV) and comparing the amount to a predetermined cutoff
CC value and thereby detecting ovarian cancer in the patient, where the
CC amount of polynucleotide hybridising to (IV) is detected preferably by
CC polymerase chain reaction (PCR). (I) comprising (III) and/or (II) is
CC useful for stimulating and/or expanding T cells specific for an ovarian
CC tumour protein comprising contacting T cells with (III) or (II). (III) is
CC useful in design and preparation of ribozyme molecules for inhibiting
CC expression of the tumour polypeptides and proteins in tumour cells; and
CC to isolate a full length gene from a suitable library e.g., a tumour cDNA
CC library using well known techniques
XX
XX Sequence 1838 BP; 563 A; 330 C; 386 G; 559 T; 0 U; 0 Other;
XX
XX Query Match 91.7%; Score 1716; DB 6; Length 1838;
XX Best Local Similarity 99.9%; Pred. No. 0;
XX Matches 1836; Conservative 0; Mismatches 0; Indels 1; Gaps 1;
XX
QY 22 GAGGCTGGCGCGCTGAGCGCGGAGCGGTTGCTGAGCCCGTTAGCGCGCGCGAG 81
DB 1 GAGGCTGGCGCGCTGAGCGCGGAGCGGTTGCTGAGCCCGTTAGCGCGCGCGAG 60
QY 82 CACGCGCGCGCATGCTCCGCTACCTGCGTCCCGCAACAGCTCTGTTGTCGAGAAC 141
DB 61 CACGCGCGCGCATGCTCCGCTACCTGCGTCCCGCAACAGCTCTGTTGTCGAGAAC 120
QY 142 GTGGCGGACGACACCGAGTGTGAAGACTTGGGCGGTGAATTTGGTGTGATAGTCTTA 201
DB 121 GTGGCGGACGACACCGAGTGTGAAGACTTGGGCGGTGAATTTGGTGTGATAGTCTTA 180
QY 202 GTTGATGTGTATGTTCCACTTATTTTACACTGCGCGTCCCAAGAGATTGCTATGTT 261
DB 181 GTTGATGTGTATGTTCCACTTATTTTACACTGCGCGTCCCAAGAGATTGCTATGTT 240
QY 262 CAATTTGAGAGATGTTGCTGATGCTGAAGACGCTTTACATTAATTTGAGAGAAATGAGATT 321
DB 241 CAATTTGAGAGATGTTGCTGATGCTGAAGACGCTTTACATTAATTTGAGAGAAATGAGATT 300
QY 322 TGTGAGCGGCGAGATTGAATATAGTTTCCGAGGGGGATCGAAAGACACCAATCAGATG 381
DB 301 TGTGAGCGGCGAGATTGAATATAGTTTCCGAGGGGGATCGAAAGACACCAATCAGATG 360

```

QY 382 AAAGCGAAGAGGAGGAAATGTGACGTTCTTACGCTATGATGATTAATGACAGATTC 441
 DB 361 AAAGCGAAGAGGAGGAAATGTGACGTTCTTACGCTATGATGATTAATGACAGATTC 420
 QY 442 AGAGCTTCTAGAACGCCAAGTTATGAAGAGGAGGAAATGAGAGGCGGCTTTGATTAC 501
 DB 421 AGAGCTTCTAGAACGCCAAGTTATGAAGAGGAGGAAATGAGAGGCGGCTTTGATTAC 480
 QY 502 AACTATAGAGATCGTATAGTCTAGAAACAGTAGACCGACTGGAAGACGCGGGTAG- 560
 DB 481 AACTATAGAGATCGTATAGTCTAGAAACAGTAGACCGACTGGAAGACGCGGGTAGC 540
 QY 561 AGAAGCCATTCCGACATGATAGACCAAACTGCGAGCTGGAATCCGACTACAGTTCTGCT 620
 DB 541 AGAAGCCATTCCGACATGATAGACCAAACTGCGAGCTGGAATCCGACTACAGTTCTGCT 600
 QY 621 TACTACACTTCAAGAAAGATCTGAAAGCGGAAAAAGAACCAAGAGGCGAGTTCAAGCG 680
 DB 601 TACTACACTTCAAGAAAGATCTGAAAGCGGAAAAAGAACCAAGAGGCGAGTTCAAGCG 660
 QY 681 ACCAAGGGTGGGTGGAAGGTGCTGAGTATGAATACTGTAAGAAATTTTGACTCTGCT 740
 DB 661 ACCAAGGGTGGGTGGAAGGTGCTGAGTATGAATACTGTAAGAAATTTTGACTCTGCT 720
 QY 741 CTGAAAAGTAAAGAAATGTTATGAAAACCTACATGGAATATTGAAGTCCCTTCAAGTT 800
 DB 721 CTGAAAAGTAAAGAAATGTTATGAAAACCTACATGGAATATTGAAGTCCCTTCAAGTT 780
 QY 801 TGAAGTATAGCATTTTGGACAAATTAAGAAATTTCACTTCTACTGTGGAACTTA 860
 DB 781 TGAAGTATAGCATTTTGGACAAATTAAGAAATTTCACTTCTACTGTGGAACTTA 840
 QY 861 TCCCTAAATATGAATAGGTTATATGATTCATGGGTACAGGTCCATAATTAATTATG 920
 DB 841 TCCCTAAATATGAATAGGTTATATGATTCATGGGTACAGGTCCATAATTAATTATG 900
 QY 921 GAAACTAGATGTCTGAATATCAAGAAAGACGCAATGCTCTTACAGTCCCTCTGTTG 980
 DB 901 GAAACTAGATGTCTGAATATCAAGAAAGACGCAATGCTCTTACAGTCCCTCTGTTG 960
 QY 981 GTCTGTCTAAACTGAATGGGTGGGAAAAAGTTAGTCCAAATTAAGTCCATTTT 1040
 DB 961 GTCTGTCTAAACTGAATGGGTGGGAAAAAGTTAGTCCAAATTAAGTCCATTTT 1020
 QY 1041 GCCATTAATGGCAATCTTGCTTTATTTTGGTCCAGTGTCTTCTGCTTAATCAT 1100
 DB 1021 GCCATTAATGGCAATCTTGCTTTATTTTGGTCCAGTGTCTTCTGCTTAATCAT 1080
 QY 1101 TTGCTTTTGGCATCTGTGTTTATTTTACTTGTACACCAATGCAAGTTTACATCTGCTT 1160
 DB 1081 TTGCTTTTGGCATCTGTGTTTATTTTACTTGTACACCAATGCAAGTTTACATCTGCTT 1140
 QY 1161 AACTACTCTTCCGAGGTAAATTCAAATTAATTGACATCCGCTTAAGGGGCCCATCT 1220
 DB 1141 AACTACTCTTCCGAGGTAAATTCAAATTAATTGACATCCGCTTAAGGGGCCCATCT 1200
 QY 1221 CTCTCACTCTTCTTCTAGTCAATTAATTCAGCAAAATTAATTGAGCCCTTACTGAGG 1280
 DB 1201 CTCTCACTCTTCTTCTAGTCAATTAATTCAGCAAAATTAATTGAGCCCTTACTGAGG 1260
 QY 1281 CAAATCATTTGTACTGGAATAATTGAGAAAAATAGATAATTCCTTATTCAGTAATATGCTTA 1340
 DB 1261 CAAATCATTTGTACTGGAATAATTGAGAAAAATAGATAATTCCTTATTCAGTAATATGCTTA 1320
 QY 1341 CTGAGACAACTAGAGTAATCACTTAAGTAATGGCCCATGTTTGTGTAAGGTGTGTA 1400
 DB 1321 CTGAGACAACTAGAGTAATCACTTAAGTAATGGCCCATGTTTGTGTAAGGTGTGTA 1380
 QY 1401 TTGATAACAATATTTTACACCAATTCGATCAATGATTAATTAGAACAAATATACGATCA 1460
 DB 1381 TTGATAACAATATTTTACACCAATTCGATCAATGATTAATTAGAACAAATATACGATCA 1440

QY 1461 AGGATTAAGTAATTTGTTGTTATCTGCCATTTAAAGATCCAGATTTGATCAATTA 1520
 DB 1441 AGGATTAAGTAATTTGTTGTTATCTGCCATTTAAAGATCCAGATTTGATCAATTA 1500
 QY 1521 TATTAATTAAGAAAAATGATTTAATCTGTAATTAATGCTGTTATTTGTCAGTGACTGTA 1580
 DB 1501 TATTAATTAAGAAAAATGATTTAATCTGTAATTAATGCTGTTATTTGTCAGTGACTGTA 1560
 QY 1581 ATATCTAGAGTTTATTAATTTGTTAATCTGCTCCACCAACATGCTAGATTA 1640
 DB 1561 ATATCTAGAGTTTATTAATTTGTTAATCTGCTCCACCAACATGCTAGATTA 1620
 QY 1641 CCCCCAAATTAAGTATTTAACTTTGCAATAGATTAAGAGACTGGGTGCTAATTAAG 1700
 DB 1621 CCCCCAAATTAAGTATTTAACTTTGCAATAGATTAAGAGACTGGGTGCTAATTAAG 1680
 QY 1701 ATTATTTTGAAGCAGACAGAGAGCTGTTATCTTAAGTATTAAGTATTTGTAATGA 1760
 DB 1681 ATTATTTTGAAGCAGACAGAGAGCTGTTATCTTAAGTATTAAGTATTTGTAATGA 1740
 QY 1761 GAAATGTTGACCAATTAATTAATTTTATAGATTTACATGATTTATAGGGACATG 1820
 DB 1741 GAAATGTTGACCAATTAATTAATTTTATAGATTTACATGATTTATAGGGACATG 1800
 QY 1821 TTCTGTATTAAGCAATTAATTAATTTTATAGATCA 1857
 DB 1801 TTCTGTATTAAGCAATTAATTAATTTTATAGATCA 1837

RESULT 4
 AAH14124
 ID AAH14124 standard; cDNA; 1836 BP.

AC AAH14124;
 XX
 XX 26-JUN-2001 (first entry)
 DT
 XX Human cDNA sequence SEQ ID NO:11319.
 DE
 XX

KW Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.

OS Homo sapiens.
 PN EP1074617-A2.
 XX
 PD 07-FEB-2001.
 PF

XX 28-JUL-2000; 2000EP-00116126.
 PR 29-AUG-1999; 99JP-00248036.
 PR 27-AUG-1999; 99JP-00300253.
 PR 11-JAN-2000; 2000JP-00118776.
 PR 02-MAY-2000; 2000JP-00183767.
 PR 09-JUN-2000; 2000JP-00241899.
 XX

PA (HELI-) HELIX RES INST.

PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
 PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
 DR WPI; 2001-318749/34.

XX Primer sets for synthesizing polynucleotides, particularly the 5602 full-
 PT length cDNAs defined in the specification, and for the detection and/or
 PT diagnosis of the abnormality of the proteins encoded by the full-length
 PT cDNAs.

PS Claim 8; SEQ ID NO 11319; 2537bp + Sequence Listing; English.

CC The present invention describes primer sets for synthesizing 5602 full-
 CC length cDNAs defined in the specification. Where a primer set comprises:
 CC (a) an oligo-dT primer and an oligonucleotide complementary to the
 CC complementary strand of a polynucleotide which comprises one of the 5602

CC nucleotide sequences defined in the specification, where the
 CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
 of an oligonucleotide comprising a sequence complementary to the
 CC complementary strand of a polynucleotide which comprises a 5'-end
 sequence and an oligonucleotide comprising a sequence complementary to a
 CC polynucleotide which comprises a 3'-end sequence, where the
 CC oligonucleotide comprises at least 15 nucleotides and the combination of the
 CC the 5'-end sequence/3'-end sequence is selected from those defined in the
 CC specification. The primer sets can be used in antisense therapy and in
 CC gene therapy. The primers are useful for synthesizing polynucleotides,
 CC particularly full-length cDNAs. The primers are also useful for the
 CC detection and/or diagnosis of the abnormality of the proteins encoded by
 CC the full-length cDNAs. The primers allow obtaining of the full-length
 CC cDNAs easily without any specialised methods. AAH3166 to AAH31628 and
 CC AAH13633 to AAH18742 represent human cDNA sequences; AAH92446 to AAH95893
 CC represent human amino acid sequences; and AAH13629 to AAH13632 represent
 CC oligonucleotides, all of which are used in the exemplification of the
 CC present invention

XX Sequence 1836 BP; 562 A; 330 C; 386 G; 558 T; 0 U; 0 Other;

Query Match 79.2%; Score 1483; DB 4; Length 1836;

Best Local Similarity 99.8%; Pred. No. 0;

Matches 1833; Conservative 0; Mismatches 1; Indels 3; Gaps 2;

22 GAGGTCGCGCCGCTGACCGCCGACGCGTTGCTGAGCCGCTTGTGCGCCGCGCGA 81
 1 GAGGTCGCGCCGCTGACCGCCGACGCGTTGCTGAGCCGCTTGTGCGCCGCGCGA 60
 82 CAGCGCGCGCGCATGTCCTGCTACCTGCGTCCGCCCAACAGCTCTGTGCTGAGAAC 141
 61 CAGCGCGCGCGCATGTCCTGCTACCTGCGTCCGCCCAACAGCTCTGTGCTGAGAAC 120
 142 GTGCGCGACGACACACAGGCTGGAAGACTTGGCGGCGTGAATTGGCTGATAGCTATA 201
 121 GTGCGCGACGACACACAGGCTGGAAGACTTGGCGGCGTGAATTGGCTGATAGCTATA 180
 202 GTTGAATGATGTTGCTTCCCTGATTTCTACACTGCGCGCTGCAAGAGATTGCTATGTT 261
 181 GTTGAATGATGTTGCTTCCCTGATTTCTACACTGCGCGCTGCAAGAGATTGCTATGTT 240
 262 CAATTGAGATGTTGCTGATGCTGAAGACGCTTACATTAATTGGAAGAAAGTGATT 321
 241 CAATTGAGATGTTGCTGATGCTGAAGACGCTTACATTAATTGGAAGAAAGTGATT 300
 322 TGTGACGCGCAGATTGAATACAGTTTGCCAGGCGGATCGAAGACACCAATCAATG 381
 301 TGTGACGCGCAGATTGAATACAGTTTGCCAGGCGGATCGAAGACACCAATCAATG 360
 382 AAAGCGAAGAGAGAGAGATGTAACGTTCTTCAAGCTATGATGATTAAGACAGATAC 441
 361 AAAGCGAAGAGAGAGAGATGTAACGTTCTTCAAGCTATGATGATTAAGACAGATAC 420
 442 AGACCTTGAAGAGCGAAGTATGAAGAGAGATCAAGAAGTCGCTCTTTGATTAC 501
 421 AGACCTTGAAGAGCGAAGTATGAAGAGAGATCAAGAAGTCGCTCTTTGATTAC 480
 502 AACCTAAGAGATGCTATGCTTGAAGACAGTACCGACTGGAAGACACGCGGTAG- 560
 481 AACCTAAGAGATGCTATGCTTGAAGACAGTACCGACTGGAAGACACGCGGTAGC 540
 561 AGAACCCATTCCGACATGATGACCAATCTGAGTGAATACCAAGTACGCTTCTGCT 620
 541 AGAACCCATTCCGACATGATGACCAATCTGAGTGAATACCAAGTACGCTTCTGCT 600
 621 TACTACACTTCAAGAAAGATCTGAAGCGGAAAAAGAACCAAGAGAGGCGACTCAAGCG 680
 601 TACTACACTTCAAGAAAGATCTGAAGCGGAAAAAGAACCAAGAGAGGCGACTCAAGCG 660
 681 ACCAAAAGGCTGGTGAAGGCTGCTGAGTATGATTAATCTGATGATTTTGAATCTGCT 740
 661 ACCAAAAGGCTGGTGAAGGCTGCTGAGTATGATTAATCTGATGATTTTGAATCTGCT 720

QY CTGAAGATTAAGAAAGATGTTATGCAAACTACATGGAATTAATTGAAGTCCCTCAAGTT 800
 DB |||||
 QY 721 CTGAAGATTAAGAAAGATGTTATGCAAACTACATGGAATTAATTGAAGTCCCTCAAGTT 780
 DB |||||
 QY 801 TGAAGTACATTTTGAAGCAAAATTAAGAAATTCACCTTGTACTGTGGAAGCTAA 860
 DB |||||
 QY 781 TGAAGTACATTTTGAAGCAAAATTAAGAAATTCACCTTGTACTGTGGAAGCTAA 840
 DB |||||
 QY 861 TCCCTAATTAAGATGATGTTATGATTCATGAGGTAACAGGTCGCAATTAATTAATG 920
 DB |||||
 QY 841 TCCCTAATTAAGATGATGTTATGATTCATGAGGTAACAGGTCGCAATTAATTAATG 900
 DB |||||
 QY 921 GAACTAGATGCTGATATCAAGAAAGACAGCATAGTCTCTTACAGTCCCTCTGTTG 980
 DB |||||
 QY 901 GAACTAGATGCTGATATCAAGAAAGACAGCATAGTCTCTTACAGTCCCTCTGTTG 960
 DB |||||
 QY 981 GTCTGTCTCAAACTGAATTTGGGTGGAAAAAGTATGCTCAATTAATTAATTAATTTT 1040
 DB |||||
 QY 961 GTCTGTCTCAAACTGAATTTGGGTGGAAAAAGTATGCTCAATTAATTAATTAATTTT 1020
 DB |||||
 QY 1041 GCCATTAATGAGCAATCTGCTGTTTATTTTGGTGGCAGGCTTTCTGTTAATGAT 1100
 DB |||||
 QY 1021 GCCATTAATGAGCAATCTGCTGTTTATTTTGGTGGCAGGCTTTCTGTTAATGAT 1080
 DB |||||
 QY 1101 TTGCTTTGTTGGCATCTGTGTTTATTTACTGTACACACATGCAAGTTTACATCTGCTT 1160
 DB |||||
 QY 1081 TTGCTTTGTTGGCATCTGTGTTTATTTACTGTACACACATGCAAGTTTACATCTGCTT 1138
 DB |||||
 QY 1161 AACTACTCTCTCCAGATTAATTCOAATTAATTTGATCTACAGTAAAGAGGCCATCT 1220
 DB |||||
 QY 1139 AACTACTCTCTCCAGATTAATTCOAATTAATTTGATCTACAGTAAAGAGGCCATCT 1198
 DB |||||
 QY 1221 CTCTGACCTCTCTCTGATGATTAATTTGCAAAATTAATTTGAGGCCCTTACGTTGAG 1280
 DB |||||
 QY 1199 CTCTGACCTCTCTCTGATGATTAATTTGCAAAATTAATTTGAGGCCCTTACGTTGAG 1258
 DB |||||
 QY 1281 CAATTCATTTGATGATTAATTTGCAAAATTAATTTGATTCATTAATGTTCTA 1340
 DB |||||
 QY 1259 CAATTCATTTGATGATTAATTTGCAAAATTAATTTGATTCATTAATGTTCTA 1318
 DB |||||
 QY 1341 CTGAGCAATCTGATGATTAATTTGCAAAATTAATTTGATTCATTAATGTTCTA 1400
 DB |||||
 QY 1319 CTGAGCAATCTGATGATTAATTTGCAAAATTAATTTGATTCATTAATGTTCTA 1378
 DB |||||
 QY 1401 TTCAATCAATTTTACACATTCGATCAATGATTAATTAAGAACCAATATGATCA 1460
 DB |||||
 QY 1379 TTCAATCAATTTTACACATTCGATCAATGATTAATTAAGAACCAATATGATCA 1438
 DB |||||
 QY 1461 AGATTAAGTAAATTTGTGTTATCTGCCATTTAAAGTATTCAGTATTTGATCAATTA 1520
 DB |||||
 QY 1439 AGATTAAGTAAATTTGTGTTATCTGCCATTTAAAGTATTCAGTATTTGATCAATTA 1498
 DB |||||
 QY 1521 TATAAATTAAGTAAATTTGATTTATCTGTTAATTAATTTGATTTGATTTGATTTGAT 1580
 DB |||||
 QY 1499 TATAAATTAAGTAAATTTGATTTATCTGTTAATTAATTTGATTTGATTTGATTTGAT 1558
 DB |||||
 QY 1581 ATATCTGAGTAAATTAATTAATTTGATTTATCTGTTAATTAATTTGATTTGATTTGAT 1640
 DB |||||
 QY 1559 ATATCTGAGTAAATTAATTAATTTGATTTATCTGTTAATTAATTTGATTTGATTTGAT 1618
 DB |||||
 QY 1641 CCCCCAAATTAAGTAAATTTGATTTATCTGTTAATTAATTTGATTTGATTTGATTTGAT 1700
 DB |||||
 QY 1619 CCCCCAAATTAAGTAAATTTGATTTATCTGTTAATTAATTTGATTTGATTTGATTTGAT 1678
 DB |||||
 QY 1701 ATTAATTTGAGGCAACAGAGGCTGTTATCTGTTAATTAATTTGATTTGATTTGATTTGAT 1760
 DB |||||
 QY 1679 ATTAATTTGAGGCAACAGAGGCTGTTATCTGTTAATTAATTTGATTTGATTTGATTTGAT 1738
 DB |||||
 QY 1761 GAAATGTTCAACCAATTAATTTGATTTATCTGTTAATTAATTTGATTTGATTTGATTTGAT 1820
 DB |||||
 QY 1739 GAAATGTTCAACCAATTAATTTGATTTATCTGTTAATTAATTTGATTTGATTTGATTTGAT 1798
 DB |||||
 QY 1821 TTCTGTATTAAGCAATTAATTAATTTGATTTATCTGTTAATTAATTTGATTTGATTTGAT 1857
 DB |||||

QY 1020 CAATATAAGTTCATTTTGGCATTAAT--GGCAATCTGCTTGTATTTTGGTG 1078
DB 1149 CAATATAAGTTCATTTTGGCATTAATGGGCAATCTGCTTGTATTTTGGTG 1208
QY 1079 CCAGTGTTCCTGCTTAATCATTTTGGTATGTCATGCTGTTATTTACTTGACACC 1138
DB 1209 CCAGTGTTCCTGCTTAATCATTTTGGTATGTCATGCTGTTATTTACTTGACACC 1268
QY 1139 ACATGCACTTTACATCTGCTTAATCACTCTCTCCAGGTAA 1180
DB 1269 ACATGCACTTTACATCTGCTTAATCACTCTCTCCAGGTAA 1310

RESULT 6
ADA52465
ID ADA52465 standard, cDNA, 2473 BP.
AC ADA52465;
XX
XX 20-NOV-2003 (first entry)
DT
XX
XX Human coding sequence, SEQ ID 33.
DE
XX
XX Cytostatic; Anti-inflammatory; Osteopathic; Neuroprotective; Nootropic;
KW Gene Therapy; human; secretory protein; membrane proteins; cancer;
KW inflammatory disease; osteoporosis; neurological disease; gene; ss.
XX
XX Homo sapiens.
OS
XX
XX EP1293569-A2.
PN
XX
XX 19-MAR-2003.
PD
XX
XX 21-MAR-2002; 2002EP-00006586.
PF
XX
XX 14-SEP-2001; 2001JP-00328381.
PR 24-JAN-2002; 2002US-0350435P.
XX
XX
XX (HELI-) HELIX RES. INST.
PA (REAS-) RES ASSOC BIOTECHNOLOGY.
XX
XX
XX Isogai T, Sugiyama T, Otsuki T, Wakamatsu A, Sato H, Ishii S,
PI Yamamoto J, Isono Y, Hio Y, Otsuka K, Nagai K, Irie R, Tamechika I,
PI Seki N, Yoshikawa T, Otsuka M, Nagahari K, Masuno Y,
XX
XX MPI: 2003-395539/38.
DR P-PSDB; ADA54104.
XX
XX
XX New polynucleotides encoding full-length polypeptides, e.g. secretory
PT and/or membrane proteins, useful for developing medicines for diseases in
PT which the gene is involved, or as target molecules for gene therapy.
XX
XX Claim 1; SEQ ID NO 33; 2055bp; English.
XX
XX The present invention relates to novel human secretory or membrane
CC proteins (ADA54072-ADA55710) and their coding sequences (ADA52433-
CC ADA54071). The coding sequences are useful in the gene therapy of
CC diseases caused by abnormalities of the proteins, e.g. cancer,
CC inflammatory diseases, osteoporosis or neurological disease.
XX
XX Sequence 2473 BP; 791 A; 395 C; 459 G; 828 T; 0 U; 0 Other;
SQ

Query Match 42.0%; Score 786; DB 7; Length 2473;
Best Local Similarity 99.9%; Pred. No. 2, 1e-300;
Matches 836; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 582 AGACCAACTGCACTGGAATACCCAGTACAGTTCCTTACTTACACTTCAAGAAAGATC 641
DB 1388 AGACCAACTGCACTGGAATACCCAGTACAGTTCCTTACTTACACTTCAAGAAAGATC 1447
QY 642 TGAAGCGGAAAAAGAACCAAGAGAGGAGAGTTCAGACCAACAAAGGAGGAGTGAAGGT 701
DB 1448 TGAAGCGGAAAAAGAACCAAGAGAGGAGAGTTCAGACCAACAAAGGAGGAGTGAAGGT 1507

QY 702 GCTGAGATGATGAATACCTGATACGAATATTTTGACTCTGCTGGAAGAGATGAATGTT 761
DB 1508 GCTGAGATGATGAATACCTGATACGAATATTTTGACTCTGCTGGAAGAGATGAATGTT 1567
QY 762 ATCGAAACTACATGAAATTAATGGAAGTCCCTCAAGTTGGAAGATGATTTAGGAC 821
DB 1568 ATCGAAACTACATGAAATTAATGGAAGTCCCTCAAGTTGGAAGATGATTTAGGAC 1627
QY 822 AAATTAAGAAATTCAACTTTGTAAGTTGGAAGATGATGATGATGATGATGATGATGAT 881
DB 1628 AAATTAAGAAATTCAACTTTGTAAGTTGGAAGATGATGATGATGATGATGATGATGAT 1687
QY 882 ATATTGATTCATGAGTGAACAGGTCATATTAATTAATTAATTAATTAATTAATTAATTAAT 941
DB 1688 ATATTGATTCATGAGTGAACAGGTCATATTAATTAATTAATTAATTAATTAATTAATTAAT 1747
QY 942 CAAGGAGACAGCCATAGTCTCTTACAGTGCCTGTTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1001
DB 1748 CAAGGAGACAGCCATAGTCTCTTACAGTGCCTGTTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1807
QY 1002 GTGGGAAAAAGGATGCTCCATATTAAGTTTCATTTTGCATTTATGGAATATCTTGC 1061
DB 1808 GTGGGAAAAAGGATGCTCCATATTAAGTTTCATTTTGCATTTATGGAATATCTTGC 1867
QY 1062 CTTGTTATTTTGGTGGCAGGTTTCTGCTTAAATCAATTTGCTTGGTGGCATCTGCT 1121
DB 1868 CTTGTTATTTTGGTGGCAGGTTTCTGCTTAAATCAATTTGCTTGGTGGCATCTGCT 1927
QY 1122 TTATTTACTTGTACACCAATGACAGTTTACATCTGCTTAACTACTCTCTCCAGGTAA 1181
DB 1928 TTATTTACTTGTACACCAATGACAGTTTACATCTGCTTAACTACTCTCTCCAGGTAA 1987
QY 1182 TTCCAAATTAATTTTACATCCAGCTAAGAGGCGCATCTCTTCTACCTCTTCCAGTGC 1241
DB 1988 TTCCAAATTAATTTTACATCCAGCTAAGAGGCGCATCTCTTCTACCTCTTCCAGTGC 2047
QY 1242 AGTATATTCAGCAATATTTATTTAGCCCTTACCTGAGGCAATCATTTGATCTGATAT 1301
DB 2048 AGTATATTCAGCAATATTTATTTAGCCCTTACCTGAGGCAATCATTTGATCTGATAT 2107
QY 1302 TGAGAAAAATAGATATTTCCCTTATTCAGTAATATGCTACTGAGCAACATCTAGGAATC 1361
DB 2108 TGAGAAAAATAGATATTTCCCTTATTCAGTAATATGCTACTGAGCAACATCTAGGAATC 2167
QY 1362 ATTACAGTATGCTCATTTGTTTGTGAGGTGTGTTATTCATACAAATATTTTAC 1418
DB 2168 ATTACAGTATGCTCATTTGTTTGTGAGGTGTGTTATTCATACAAATATTTTAC 2224

RESULT 7
ABV25632
ID ABV25632 standard, cDNA, 1382 BP.
XX
XX
XX ABV25632;
AC
XX
XX 16-SEP-2002 (first entry)
DT
XX
XX Human prostate expression marker CDNA 25623.
DE
XX
XX Human; prostate cancer; cytostatic; carcinogen; pharmacodynamic marker;
KW pharmacogenomic marker; gene; ss.
XX
XX Homo sapiens.
OS
XX
XX WO200160860-A2.
PN
XX
XX 23-AUG-2001.
PD
XX
XX 20-FEB-2001; 2001WO-US005171.
PF
XX
XX 17-FEB-2000; 2000US-0183319P.
PR 16-MAR-2000; 2000US-0189862P.

25-MAY-2000; 2000US-0207454P.
PR 09-JUN-2000; 2000US-0211314P.
PR 18-JUN-2000; 2000US-0219007P.
PR 13-DEC-2000; 2000US-0255281P.
XX
XX
XX (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
PI Schlegel R, Endege WO, Monahan JE;
XX WPI; 2001-662795/76.
XX
XX Novel isolated nucleic acid molecule associated with cancerous state of
PT prostate cells and correlating with presence of prostate cancer, useful
PT for detecting presence of prostate cancer, stage of prostate cancer.
XX
XX Claim 1; Page 5128-5129; 11750pp; English.
XX
XX The invention relates to an isolated nucleic acid molecule (I) comprising
CC a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the
CC specification or its complement. (I) is useful for: (a) assessing whether
CC a patient is afflicted with prostate cancer; (b) monitoring the
CC progression of prostate cancer in a patient; (c) assessing the efficacy
CC of a test compound to inhibit prostate cancer in a patient; (d) assessing
CC the efficacy of a therapy for inhibiting prostate cancer in a patient;
CC (e) selecting a composition for inhibiting prostate cancer in a patient;
CC (f) assessing the prostate cell carcinogenic potential of a compound; (g)
CC determining whether prostate cancer has metastasized in a patient; (h)
CC assessing the aggressiveness or indolence of prostate cancer in a patient
CC ; (I) is also useful as a pharmacodynamic or pharmacogenomic marker
XX
SQ Sequence 1382 BP; 462 A; 282 C; 285 G; 340 T; 0 U; 13 Other;
Query Match 37.9%; Score 709; DB 5; Length 1382;
Best Local Similarity 100.0%; Pred. No. 5,5e-270;
Matches 709; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 561 AGAAGCCATTCGCAATGATGACCAACTGCGCTGGAATACCCAGTACGTTCTGCT 620
DB 230 AGAAGCCATTCGCAATGATGACCAACTGCGCTGGAATACCCAGTACGTTCTGCT 289
QY 621 TACACACCTCCAAAGAAAGATCTGAAAGCGGAAAGAAAGAAAGCGGCGTCAAGCG 680
DB 290 TACACACCTCCAAAGAAAGATCTGAAAGCGGAAAGAAAGAAAGCGGCGTCAAGCG 349
QY 681 ACCAAGGGTGGTGGAGGCTGCGATGATGATGATGATGATGATGATGATGATGATGAT 740
DB 350 ACCAAGGGTGGTGGAGGCTGCGATGATGATGATGATGATGATGATGATGATGATGAT 409
QY 741 CTGAAAAGATGAAAGATGTTATGCAAACTACATGAAATATGAAAGTCCCTCAAGTT 800
DB 410 CTGAAAAGATGAAAGATGTTATGCAAACTACATGAAATATGAAAGTCCCTCAAGTT 469
QY 801 TGAAGTAAGCATTTTGGCAAAATGAAAGAAATGCACTTCTACTTGGGAACTAA 860
DB 470 TGAAGTAAGCATTTTGGCAAAATGAAAGAAATGCACTTCTACTTGGGAACTAA 529
QY 861 TCCCTAAATGATAGATTTATGATGATGATGATGATGATGATGATGATGATGATGAT 920
DB 530 TCCCTAAATGATAGATTTATGATGATGATGATGATGATGATGATGATGATGATGAT 589
QY 921 GAAATCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 980
DB 590 GAAATCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 649
QY 981 GTCTGCTCAAACTGAAATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1040
DB 650 GTCTGCTCAAACTGAAATGATGATGATGATGATGATGATGATGATGATGATGATGAT 709
QY 1041 GCCATTTATGCGCAAAATCTTGCCTTTGTTATTTTGGTGGCAAGTCTTTTCTTAATCAT 1100
DB 710 GCCATTTATGCGCAAAATCTTGCCTTTGTTATTTTGGTGGCAAGTCTTTTCTTAATCAT 769
QY 1101 TTGCTTTGTTGGCATCTGCTTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 1160

DB 770 TTGCTTTGTTGGCATCTGCTTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 829
QY 1161 AACACGCTCTCCAGGTAATTTCCATTTATTTGATGATGATGATGATGATGATGATGAT 1220
DB 830 AACACGCTCTCCAGGTAATTTCCATTTATTTGATGATGATGATGATGATGATGATGAT 889
QY 1221 CTTCGACCTCTCTCTAGTATGATGATGATGATGATGATGATGATGATGATGATGAT 1269
DB 890 CTTCGACCTCTCTCTAGTATGATGATGATGATGATGATGATGATGATGATGATGAT 938
RESULT 8
ID AAH14349
AAH14349 standard; cDNA; 2247 BP.
XX
XX AAH14349;
AC
XX
XX 26-JUN-2001 (first entry)
DT
XX
XX Human cDNA sequence SEQ ID NO:11739.
DE
XX
XX Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
KM
XX
XX Homo sapiens.
OS
XX
XX EP1074617-A2.
PN
XX
XX 07-FEB-2001.
PD
XX
XX 28-JUN-2000; 2000EP-00116126.
PF
XX
XX 29-JUN-1999; 99JP-00248036.
PR 27-AUG-1999; 99JP-00300253.
PR 11-JUN-2000; 2000JP-00118776.
PR 02-MAY-2000; 2000JP-00183767.
PR 09-JUN-2000; 2000JP-00241899.
XX
XX (HELI-) HELIX RES INST.
PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saico K, Yamamoto J;
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX
XX WPI; 2001-318749/34.
DR
XX
XX Primer sets for synthesizing polynucleotides, particularly the 5602 full-
PT length cDNAs defined in the specification, and for the detection and/or
PT diagnosis of the abnormality of the proteins encoded by the full-length
PT cDNAs.
PT
XX
XX Claim 8; SEQ ID NO 11739; 2537pp + Sequence Listing; English.
XX
XX The present invention describes primer sets for synthesizing 5602 full-
CC length cDNAs defined in the specification. Where a primer set comprises:
CC (a) an oligo-dT primer and an oligonucleotide complementary to the
CC complementary strand of a polynucleotide which comprises one of the 5602
CC nucleotide sequences defined in the specification, where the
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC of an oligonucleotide comprising a sequence complementary to the
CC complementary strand of a polynucleotide which comprises a 5'-end
CC sequence and an oligonucleotide comprising a sequence complementary to a
CC polynucleotide which comprises a 3'-end sequence, where the
CC oligonucleotide comprises at least 15 nucleotides and the combination
CC of the 5'-end sequence/3'-end sequence is selected from those defined in the
CC specification. The primer sets can be used in antisense therapy and in
CC gene therapy. The primers are useful for synthesizing polynucleotides,
CC particularly full-length cDNAs. The primers are also useful for the
CC detection and/or diagnosis of the abnormality of the proteins encoded by
CC the full-length cDNAs. The primers allow obtaining of the full-length
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC AAH13633 to AAH18742 represent human cDNA sequences; AAH92446 to AAH95893
CC represent human amino acid sequences; and AAH13629 to AAH13632 represent
CC oligonucleotides, all of which are used in the exemplification of the

CC present invention

XX Sequence 2247 BP, 669 A, 348 C, 389 G, 841 T; 0 U; 0 Other;

XX Query Match 36.0%; Score 674; DB 4; Length 2247;

XX Best Local Similarity 99.9%; Pred. No. 3,4e-256;

XX Matches 724; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1133 TACACCAATGAGATTGATACATCTGCTTACACTACCTCCAGGTAATTCATTTA 1192

Db 1 TACACCAATGAGATTGATACATCTGCTTACACTACCTCCAGGTAATTCATTTA 60

QY 1193 TTGACATCCAGCTAGAGAGGGCCCATCTCTGACCTCTTCCAGTACATATTAG 1252

Db 61 TTGACATCCAGCTAGAGAGGGCCCATCTCTGACCTCTTCCAGTACATATTAG 120

QY 1253 CAATATTTATTTAGACCCCTTACTGCTGGGCAATTCATTTGATGATTAATGAAAAATA 1312

Db 121 CAATATTTATTTAGACCCCTTACTGCTGGGCAATTCATTTGATGATTAATGAAAAATA 180

QY 1313 GATATTTCCCTATTCAGTAAATGCTCTAGTGCACATCTGATGATTAATGAGAAATTA 1372

Db 181 GATATTTCCCTATTCAGTAAATGCTCTAGTGCACATCTGATGATTAATGAGAAATTA 240

QY 1373 GCTCATTTGTTTGTGAGTGTGTATTCATTAACAATTTTACACCATTCGATGAA 1432

Db 241 GCTCATTTGTTTGTGAGTGTGTATTCATTAACAATTTTACACCATTCGATGAA 300

QY 1433 TGTAAATATAGAACCAATATATGATCAAGATTAAGTAATGTGTGTTATCTGCCATT 1492

Db 301 TGTAAATATAGAACCAATATATGATCAAGATTAAGTAATGTGTGTTATCTGCCATT 360

QY 1493 AAAAGATCCCACTATTTGATCACTATTAATTAATGAAATTAATGATTAATCTGTA 1552

Db 361 AAAAGATCCCACTATTTGATCACTATTAATTAATGAAATTAATGATTAATCTGTA 420

QY 1553 TAACTGTTTATTTGAGAGTGTATATATATATATATATATATATATATATATAT 1612

Db 421 TAACTGTTTATTTGAGAGTGTATATATATATATATATATATATATATATATAT 480

QY 1613 GCTTACCAACACATGCTAGATATACCCCAAAATTAATTAATTTACTTTGATTAG 1672

Db 481 GCTTACCAACACATGCTAGATATACCCCAAAATTAATTAATTTACTTTGATTAG 540

QY 1673 TATTAAGGAGACTGGGCTATTAATTAATTTTGAAGGAGAGAGAGGCTGTTATCC 1732

Db 541 TATTAAGGAGACTGGGCTATTAATTAATTTTGAAGGAGAGAGAGGCTGTTATCC 600

QY 1733 TAACTGATTTAGTATGCTGTATTAATGAGAAATGTTCAACAAATTAATTTTATGTA 1792

Db 601 TAACTGATTTAGTATGCTGTATTAATGAGAAATGTTCAACAAATTAATTTTATGTA 660

QY 1793 TTTAATGATGATTTTATATGAGGAGATGTTCTGTATATGCGAATTAATTAATTTTATG 1852

Db 661 TTTAATGATGATTTTATATGAGGAGATGTTCTGTATATGCGAATTAATTAATTTTATG 720

QY 1853 TATCA 1857

Db 721 TATCA 725

RESULT 9

AAH06844 ID AAH06844 standard; cDNA; 702 BP.

XX AAH06844;

XX 26-JUN-2001 (first entry)

XX Human cDNA clone (5'-primer) SEQ ID NO:3679.

XX Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.

OS Homo sapiens.

XX EPI074617-A2.

XX 07-FEB-2001.

XX 28-JUL-2000; 2000EP-00116126.

XX 29-JUL-1999; 99JP-00248036.

XX 27-AUG-1999; 99JP-00300253.

XX 11-JAN-2000; 2000JP-00118776.

XX 02-MAY-2000; 2000JP-00183767.

XX 09-JUN-2000; 2000JP-00241899.

XX (HELI-) HELIX RES INST.

XX Ota T, Isegai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;

XX Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;

XX WPI; 2001-318749/34.

PS Claim 1; SEQ ID NO 3679; 2537bp + Sequence Listing; English.

XX The present invention describes primer sets for synthesizing 5602 full-length cDNAs defined in the specification. Where a primer set comprises:

CC (a) an oligo-dT primer and an oligonucleotide complementary to the

CC complementary strand of a polynucleotide which comprises one of the 5602

CC nucleotide sequences defined in the specification, where the

CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination

CC of an oligonucleotide comprising a sequence complementary to the

CC complementary strand of a polynucleotide which comprises a 5'-end

CC sequence and an oligonucleotide comprising a sequence complementary to a

CC polynucleotide which comprises a 3'-end sequence, where the

CC oligonucleotide comprises at least 15 nucleotides and the combination of

CC the 5'-end sequence/3'-end sequence is selected from those defined in the

CC specification. The primer sets can be used in antisense therapy and in

CC gene therapy. The primers are useful for synthesizing polynucleotides,

CC particularly full-length cDNAs. The primers are also useful for the

CC detection and/or diagnosis of the abnormality of the proteins encoded by

CC the full-length cDNAs. The primers allow obtaining of the full-length

CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and

CC AAH13633 to AAH16742 represent human cDNA sequences; AAH92446 to AAH95893

CC represent human amino acid sequences; and AAH13629 to AAH13632 represent

CC oligonucleotides, all of which are used in the exemplification of the

CC present invention

CC Sequence 702 BP; 201 A; 159 C; 173 G; 166 T; 0 U; 3 Other;

QY Query Match 27.2%; Score 509; DB 4; Length 702;

Db Best Local Similarity 100.0%; Pred. No. 5.1e-191; Indels 0; Gaps 0; Matches 509; Conservative 0; Mismatches 0;

QY 22 GAGGCTGGCGCGCTGAGCGCGGAGCGGTTGCTGAGCCCGTTAGTGCAGCGCGGAGAA 81

Db 1 GAGGCTGGCGCGCTGAGCGCGGAGCGGTTGCTGAGCCCGTTAGTGCAGCGCGGAGAA 60

QY 82 CAGCGCGCGCGCATGCTCCGCTACCTGCGTCCGCCCAACAGCGTCTGTGTGTCAGGAAC 141

Db 61 CAGCGCGCGCGCATGCTCCGCTACCTGCGTCCGCCCAACAGCGTCTGTGTGTCAGGAAC 120

QY 142 GTGGCCGACGACACGAGTCTGAAGACTTGGCGGCGTGAATTTGTCGTATAGTCTATA 201

Db 121 GTGGCCGACGACACGAGTCTGAAGACTTGGCGGCGTGAATTTGTCGTATAGTCTATA 180

QY 202 GTTGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 261

Db 181 GTTGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 240

QY 262 CAATTGGAGATGTTCTGATGCTGAAGACGCTTTACATTAATTGGACAGAAAGTGATT 321
 Db 241 CAATTGGAGATGTTCTGATGCTGAAGACGCTTTACATTAATTGGACAGAAAGTGATT 300
 QY 322 TGTGGACGGCAGATTGAATACAGTTTCCGAGGGGGATGCAAAAGACCAAAATCAGATG 381
 Db 301 TGTGGACGGCAGATTGAATACAGTTTCCGAGGGGGATGCAAAAGACCAAAATCAGATG 360
 QY 382 AAAGCCAAAGAGGAGGAGATGTGTACAGTTCTTACAGCTATGTGATTTATGACAGATAC 441
 Db 361 AAAGCCAAAGAGGAGGAGATGTGTACAGTTCTTACAGCTATGTGATTTATGACAGATAC 420
 QY 442 AGAGCTTCTAGAACCCGAAAGTTATGAAAGAGAGATCAGAAGTCGGTCTTTTGATTAC 501
 Db 421 AGAGCTTCTAGAACCCGAAAGTTATGAAAGAGAGATCAGAAGTCGGTCTTTTGATTAC 480
 QY 502 AACTATAGAAAGATCGTATAGTCTTAGAA 530
 Db 481 AACTATAGAAAGATCGTATAGTCTTAGAA 509

RESULT 10

AAH14554
 ID AAH14554 standard; cDNA; 2918 BP.

XX AAH14554;
 AC
 XX 26-JUN-2001 (first entry)
 DT
 XX

DE Human cDNA sequence SEQ ID NO:12123.

XX Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.

OS Homo sapiens.

XX EP1074617-A2.

XX 07-FEB-2001.

XX 28-JUL-2000; 2000EP-00116126.

XX 29-JUL-1999; 99JP-00248036.

XX 27-AUG-1999; 99JP-00300253.

XX 11-JAN-2000; 2000JP-00118776.

XX 02-MAY-2000; 2000JP-00183767.

XX 09-JUN-2000; 2000JP-00241899.

XX (HELI-) HELIX RES INST.

XX Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
 PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
 XX WPI; 2001-318749/34.

PT Primer sets for synthesizing polynucleotides, particularly the 5602 full-length cDNAs defined in the specification, and for the detection and/or diagnosis of the abnormality of the proteins encoded by the full-length cDNAs.

XX Claim 8; SEQ ID NO 12123; 2537bp + Sequence listing; English.

XX The present invention describes primer sets for synthesizing 5602 full-length cDNAs defined in the specification. Where a primer set comprises: (a) an oligo-dT primer and an oligonucleotide complementary to the complementary strand of a polynucleotide which comprises one of the 5602 oligonucleotide sequences defined in the specification, where the oligonucleotide comprises at least 15 nucleotides; or (b) a combination of an oligonucleotide comprising a sequence complementary to the complementary strand of a polynucleotide which comprises a 5'-end sequence and an oligonucleotide comprising a sequence complementary to a polynucleotide which comprises a 3'-end sequence, where the oligonucleotide comprises at least 15 nucleotides and the combination of the 5'-end sequence/3'-end sequence is selected from those defined in the

CC specification. The primer sets can be used in antisense therapy and in gene therapy. The primers are useful for synthesizing polynucleotides, particularly full-length cDNAs. The primers are also useful for the detection and/or diagnosis of the abnormality of the proteins encoded by the full-length cDNAs. The primers allow obtaining of the full-length cDNAs easily without any specialised methods. AAH03166 to AAH13628 and AAH13633 to AAH18742 represent human cDNA sequences; AAH92446 to AAH95893 represent human amino acid sequences; and AAH13629 to AAH13632 represent oligonucleotides, all of which are used in the exemplification of the present invention

XX Sequence 2918 BP; 889 A; 525 C; 614 G; 890 T; 0 U; 0 Other;

Query Match 27.2%; Score 509; DB 4; Length 2918;

Best Local Similarity 100.0%; Pred. No. 4,2e-191;

Matches 509; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 22 GAGGCTCGCCGCGCTGAGCCGCGGACGGTTGTCTGAGCCCGTTAGTGCAGCCGCGGAG 81
 Db 1 GAGGCTCGCCGCGCTGAGCCGCGGACGGTTGTCTGAGCCCGTTAGTGCAGCCGCGGAG 60

QY 82 CAGCGCGCGCGCAGTGTCCGCTACCTGCGTCCCGCAACAGCTCTGTGTAGAGAC 141
 Db 61 CAGCGCGCGCGCAGTGTCCGCTACCTGCGTCCCGCAACAGCTCTGTGTAGAGAC 120

QY 142 GTGGCCGACGACACAGGTCTGAAGACTTGGCGGCGTGAATTTGGTTCGTTAGCTTATA 201
 Db 121 GTGGCCGACGACACAGGTCTGAAGACTTGGCGGCGTGAATTTGGTTCGTTAGCTTATA 180

QY 202 GTTGATGTGTATGTTTCACCTGATTTCTACACTGCGCGTCCAGAGAGATTTGCTATGTT 261
 Db 181 GTTGATGTGTATGTTTCACCTGATTTCTACACTGCGCGTCCAGAGAGATTTGCTATGTT 240

QY 262 CAATTGAGATGTTGCTGAGTGTGAGAGAGCTTTATATATTTGGACAGAAATGAGATT 321
 Db 241 CAATTGAGATGTTGCTGAGTGTGAGAGAGCTTTATATATTTGGACAGAAATGAGATT 300

QY 322 TGTGGACGGCAGATTGAATACAGTTTCCGAGGGGATGCAAAAGACCAAAATCAGATG 381
 Db 301 TGTGGACGGCAGATTGAATACAGTTTCCGAGGGGATGCAAAAGACCAAAATCAGATG 360

QY 382 AAAGCCAAAGAGGAGGAGAAATGTGTACAGTTCTTACAGCTATGTGATTTATGACAGATAC 441
 Db 361 AAAGCCAAAGAGGAGGAGAAATGTGTACAGTTCTTACAGCTATGTGATTTATGACAGATAC 420

QY 442 AGAGCTTCTAGAACCCGAAAGTTATGAAAGAGAGATCAGAAGTCGGTCTTTTGATTAC 501
 Db 421 AGAGCTTCTAGAACCCGAAAGTTATGAAAGAGAGATCAGAAGTCGGTCTTTTGATTAC 480

QY 502 AACTATAGAAAGATCGTATAGTCTTAGAA 530
 Db 481 AACTATAGAAAGATCGTATAGTCTTAGAA 509

RESULT 11

ABL87933/c

ID ABL87933 standard; DNA; 550 BP.

XX ABL87933;
 AC
 XX 17-MAY-2002 (first entry)
 DT
 XX

DE Human ovarian cancer related DNA clone SEQ ID NO:10911.

XX Human; ovarian cancer; ovarian tumour; cytostatic; gene; ds.

XX Homo sapiens.

XX WO200192581-A2.

XX 06-DEC-2001.

XX 29-MAY-2001; 2001WO-US017756.

OY	126	TCGTGTTGTGACGAGACGTGGCCGACGACACACAGGTCTGAAGACTTGGCGGCTGAATTTGG	185
Db	144	TCGTGTTGTGACGAGACGTGGCCGACGACACAGGTCTGAAGACTTGGCGGCTGAATTTGG	203
OY	186	TCGTTATGAGTCCCTATAGTGTGAAGTGTAAGTTCACATTGATTTTCTACATTCGCGCGTCCAG	245
Db	204	TCGTTATGAGTCCCTATAGTGTGAAGTGTAAGTTCACATTGATTTTCTACATTCGCGCGTCCAG	263
OY	246	AGGATTTGCTTATGTTCAATTTGAGAGATGTTCCGTATGGTGAAGACGCTTTACATTAATTT	305
Db	264	AGGATTTGCTTATGTTCAATTTGAGAGATGTTCCGTATGGTGAAGACGCTTTACATTAATTT	323
OY	306	GGAACAGAAAGTGAATTTGTGGAACGGACATTTGAAAATACATTTGGCCACGAGGAGATCGAAA	365
Db	324	GGAACAGAAAGTGAATTTGTGGAACGGACATTTGAAAATACATTTGGCCACGAGGAGATCGAAA	383
OY	366	GACACCCAAATCAATGATAAAGCCACGAGAAAGGAGAGATGTGTACATTTCTTACGCTATGA	425
Db	384	GACACCCAAATCAATGATAAAGCCACGAGAAAGGAGAGATGTGTACATTTCTTACGCTATGA	443
OY	426	TGATTTATGACAGATTACAGACGTTCTAGAACCGAAGTTATGAAAAGAGAGATCAGAG	485
Db	444	TGATTTATGACAGATTACAGACGTTCTAGAACCGAAGTTATGAAAAGAGAGATCAGAG	503
OY	486	TCGGTCTTTTGAATTAACAATTAAGAAGATCGTATAGTCCCTAGAAAACAGTAGACGACTGG	545
Db	504	TCGGTCTTTTGAATTAACAATTAAGAAGATCGTATAGTCCCTAGAAAACAGTAGACGACTGG	563
OY	546	AAGACCCAGGCGCTGAG	560
Db	564	AAGACCCAGGCGCTGAG	578
RESULT 13			
AA	64869		
ID	AA64869	standard; cDNA; 716 BP.	
AC	AA64869;		
XX			
DT	13-FEB-2002	(first entry)	
DE	DNA encoding novel human diagnostic protein #673.		
XX			
KW	Human; chromosome mapping; gene mapping; gene therapy; forensic;		
XX			
KW	food supplement; medical imaging; diagnostic; genetic disorder; ss.		
XX			
OS	Homo sapiens.		
XX			
FN	W0200175067-A2.		
XX			
PD	11-OCT-2001.		
XX			
PF	30-MAR-2001; 2001MO-US008631.		
XX			
PR	31-MAR-2000; 2000US-00540217.		
XX			
PR	23-AUG-2000; 2000US-00649167.		
XX			
PA	(HYSE-) HYSEQ INC.		
XX			
PI	Drmanac RT, Liu C, Tang YF,		
XX			
DR	WPI; 2001-639362/73.		
XX			
DR	P-PSDB; ABG00682.		
XX			
XX	New isolated polynucleotide and encoded polypeptides, useful in		
PT	diagnostics, forensics, gene mapping, identification of mutations		
PT	responsible for genetic disorders or other traits and to assess		
PT	biodiversity.		
XX			
ES	Claim 1; SEQ ID NO 673; 103bp; English.		
XX			
CC	The invention relates to isolated polynucleotide (I) and polypeptide (II)		
CC	sequences. (I) is useful as hybridisation probes, polymerase chain		

	CC	reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
	CC	and in recombinant production of (II). The polynucleotides are also used
	CC	in diagnostics as expressed sequence tags for identifying expressed
	CC	genes. (I) is useful in gene therapy techniques to restore normal
	CC	activity of (II) or to treat disease states involving (II). (II) is
	CC	useful for generating antibodies against it, detecting or quantitating a
	CC	polypeptide in tissue, as molecular weight markers and as a food
	CC	supplement. (II) and its binding partners are useful in medical imaging
	CC	of sites expressing (II). (I) and (II) are useful for treating disorders
	CC	involving aberrant protein expression or biological activity. The
	CC	polypeptide and polynucleotide sequences have applications in
	CC	diagnostics, forensics, gene mapping, identification of mutations
	CC	responsible for genetic disorders or other traits to assess biodiversity
	CC	and to produce other types of data and products dependent on DNA and
	CC	amino acid sequences. AAS64197-AAS94564 represent novel human diagnostic
	CC	coding sequences of the invention. Note: The sequence data for this
	CC	patent did not appear in the printed specification, but was obtained in
	CC	electronic format directly from WIPO at
	CC	ftp.wipo.int/pub/published_pct_sequences
SQ	xx	
	Sequence	716 BP; 194 A; 163 C; 177 G; 182 T; 0 U; 0 Other;
Query Match	23.2%;	Score 435; DB 5; Length 716;
Best Local Similarity	100.0%;	Pred. No. 8e-162;
Matches 435;	Conservative 0;	Mismatches 0; Indels 0; Gaps 0;
OY	126	TCTGTTGTGCAGGAACGTGGCGCAGCACCAAGSTTGGAGACTTGCGCGGTGAATTTGG 185
Dd	144	TCGTGTGTCGACGAACTGGCGCAGCACCAAGGTGGAAGACTTGCGCGGTGAATTTGG 203
OY	186	TCGTTAATGTCCTTAAGTGTGATGCTGTGATTGCCACTTGATTTCTAACCTCGCGTCCAAG 245
Dd	204	TCGTTAATGTCCTTAAGTGTGATGCTGTGATTGCCACTTGATTTCTAACCTCGCGTCCAAG 263
OY	246	AGAATTTGCTTAATGTCOAATTTGAGAAATGTTGCTGATGCTGAAGAAGCTTTACAATAATT 305
Dd	264	AGGATTTGCTTAATGTCOAATTTGAGAAATGTTGCTGATGCTGAAGAAGCTTTACAATAATT 323
OY	306	GGAACAAAGTAGAATTTGTGAGCGGCGCATTTGAAAATCAGTTTTCCGACGGGGATCCGAA 365
Dd	324	GGAACAAAGTAGAATTTGTGAGCGGCGCATTTGAAAATCAGTTTTCCGACGGGGATCCGAA 383
OY	366	GACACCAAATCAGATGGAAGCGCAAGGAAGGAGGAATGTGTACAGTTCCTTCAACCTTAGA 425
Dd	384	GACACCAAATCAGATGGAAGCGCAAGGAAGGAGGAATGTGTACAGTTCCTTCAACCTTAGA 443
OY	426	TGATTTATGACAGATACAGACGTTCTTAGAAGCCGAAGTTATGAAAAAGAGAGATCAAGAAG 485
Dd	444	TGATTTATGACAGATACAGACGTTCTTAGAAGCCGAAGTTATGAAAAAGAGAGATCAAGAAG 503
OY	486	TCGGTCTTTTGAATTAACAATTTAGAAATGCTTAAGTCTTGAAGAAACGATACCGACTGG 545
Dd	504	TCGGTCTTTTGAATTAACAATTTAGAAATGCTTAAGTCTTGAAGAAACGATACCGACTGG 563
OY	546	AAGACCAAGCGGCGTAG 560
Dd	564	AAGACCAAGCGGCGTAG 578
RESULT 14		
ID	AA196798	standard; cDNA; 737 BP.
XX	AA196798;	
XX	DT	13-NOV-2001 (first entry)
DE	Human neuroblastoma expressed polynucleotide SEQ ID NO 2873.	
XX	Human; neuroblastoma; malignancy; cancer; tumour marker; N-myc; TrkA; ss.	
OS	Homo sapiens.	
XX		

```
EN WO20016719-A1.
XX
XX 13-SEP-2001.
XX
XX 02-MAR-2001; 2001WO-JP001629.
XX
XX 07-MAR-2000; 2000JP-00159195.
XX
XX (CHIB-) CHIBA PREFECTURE.
XX PA (HISM) HISAMITSU PHARM CO LTD.
XX PI Nakagawara A;
XX DR WPI; 2001-565584/63.
XX
XX Nucleic acids originating in gene expressed in human neuroblastoma,
XX PT useful as probe or primer in diagnosing prognosis of human neuroblastoma,
XX PT malignancy and susceptibility indicator or tumor marker for anti-cancer
XX PT agents.
XX
XX Claim 1; Page 2100; 2979pp; Japanese.
XX
XX The invention relates to novel genes (AA193926-AA197963) expressed in
XX CC human neuroblastoma. The nucleic acids are applicable as a probe or
XX CC primer in diagnosing the prognosis of human neuroblastoma, malignancy and
XX CC susceptibility indicators or tumour markers for anti-cancer agents. The
XX CC gene information for diagnosing prognosis is related to factors similar
XX CC to that for N-myc and Yrka genes
XX
XX Sequence 737 BP; 213 A; 121 C; 128 G; 255 T; 0 U; 20 Other;
SQ
Query Match 23.0%; Score 430; DB 4; Length 737;
Best Local Similarity 99.8%; Pred. No. 7.5e-160;
Matches 480; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1069 TATTTGGTGGCAGTGTTCCTTATTCATTTGCTTGTGGCATCTGTTATTTA 1128
DB 48 TATTTGGTGGCAGTGTTCCTTATTCATTTGCTTGTGGCATCTGTTATTTA 107
QY 1129 CTGTGACCCACATGACATGCTTACTCTCTCCAGGTAATTCAT 1188
DB 108 CTGTGACCCACATGACATGCTTACTCTCTCCAGGTAATTCAT 167
QY 1189 TATATTTGACATCCAGCTAGAGGCGCCATCTCTACCTCTTCTAGCAGTATAT 1248
DB 168 TATATTTGACATCCAGCTAGAGGCGCCATCTCTACCTCTTCTAGCAGTATAT 227
QY 1249 TCAGCAATATTTATTTGAGCCCTTACTGTGGCAATCATTTGATGATTAAGAAA 1308
DB 228 TCAGCAATATTTATTTGAGCCCTTACTGTGGCAATCATTTGATGATTAAGAAA 287
QY 1309 AATAGATATTTCCCTTATTCAGTAATGTCTACTGAGCAATCTAGTGAATCTTAAG 1368
DB 288 AATAGATATTTCCCTTATTCAGTAATGTCTACTGAGCAATCTAGTGAATCTTAAG 347
QY 1369 TATGGCCATCTGTTGTTGAGGTGTATTTATTAACAATATTTTACACCATTCGTA 1428
DB 348 TATGGCCATCTGTTGTTGAGGTGTATTTATTAACAATATTTTACACCATTCGTA 407
QY 1429 TCAATGTATTTATAGAACACATATATAGATCAAGATAGTAATGTGTGTTATCTGCC 1488
DB 408 TCAATGTATTTATAGAACACATATATAGATCAAGATAGTAATGTGTGTTATCTGCC 467
QY 1489 ATTTAAAGTATCCGATTTTGTATGATCATATATTAATAAATAAATGATTTAACT 1548
DB 468 ATTTAAAGTATCCGATTTTGTATGATCATATATTAATAAATAAATAAATGATTTAACT 527
QY 1549 G 1549
DB 528 G 528
RESULT 15
```

```
ABV24425
ID ABV24425 standard; cDNA; 2407 BP.
XX
XX AC ABV24425;
XX
XX DT 16-SEP-2002 (first entry)
XX
XX DE Human prostate expression marker cDNA 24416.
XX
XX KW Human; prostate cancer; cytostatic; carcinogen; pharmacodynamic marker;
XX KW pharmacogenomic marker; gene; ss.
XX
XX OS Homo sapiens.
XX
XX PN WO200160860-A2.
XX
XX PD 23-AUG-2001.
XX
XX PF 20-FEB-2001; 2001WO-US005171.
XX
XX PR 17-FEB-2000; 2000US-0183319P.
XX PR 16-MAR-2000; 2000US-0189862P.
XX PR 25-MAY-2000; 2000US-0207454P.
XX PR 09-JUN-2000; 2000US-0211314P.
XX PR 18-JUL-2000; 2000US-0219007P.
XX PR 13-DEC-2000; 2000US-0255281P.
XX
XX (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX
XX PI Schlegel R, Endege WO, Monahan JE;
XX
XX DR WPI; 2001-662795/76.
XX
XX PT Novel isolated nucleic acid molecule associated with cancerous state of
XX PT prostate cells and correlating with presence of prostate cancer; useful
XX PT for detecting presence of prostate cancer, stage of prostate cancer.
XX
XX PS Claim 1; Page 4603; 11750pp; English.
XX
XX CC The invention relates to an isolated nucleic acid molecule (I) comprising
XX CC a nucleotide sequence given in Tables 1-9 (ABV00010-ABV6213) of the
XX CC specification or its complement. (I) is useful for: (a) assessing whether
XX CC a patient is afflicted with prostate cancer; (b) monitoring the
XX CC progression of prostate cancer in a patient; (c) assessing the efficacy
XX CC of a test compound to inhibit prostate cancer in a patient; (d) assessing
XX CC the efficacy of a therapy for inhibiting prostate cancer in a patient;
XX CC (e) selecting a composition for inhibiting prostate cancer in a patient;
XX CC (f) assessing the prostate cell carcinogenic potential of a compound; (g)
XX CC determining whether prostate cancer has metastasized in a patient; (h)
XX CC assessing the aggressiveness or indolence of prostate cancer in a patient
XX CC ; (I) is also useful as a pharmacodynamic or pharmacogenomic marker
XX
XX SQ Sequence 2407 BP; 781 A; 402 C; 526 G; 694 T; 0 U; 4 Other;
Query Match 22.8%; Score 426; DB 5; Length 2407;
Best Local Similarity 100.0%; Pred. No. 2.4e-158;
Matches 426; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 582 AGACCAATCTGACGCTGGAATTAACCAATGACGTTGCTTACTACCTTCAAGAAAGATC 641
DB 1671 AGACCAATCTGACGCTGGAATTAACCAATGACGTTGCTTACTACCTTCAAGAAAGATC 1730
QY 642 TGAAGCGGAAAGAAAGAACCAAGAGGCGAGTTCAAGGACCAAGGCGGTGGAAGT 701
DB 1731 TGAAGCGGAAAGAAAGAACCAAGAGGCGAGTTCAAGGACCAAGGCGGTGGAAGT 1790
QY 702 GCTGCAATGAAATATCTGATCAATATTTGACTGCTGCTGAAAGATPAAAGATGTT 761
DB 1791 GCTGCAATGAAATATCTGATCAATATTTGACTGCTGCTGAAAGATPAAAGATGTT 1850
QY 762 ATCGAAATCTACATGGAATTAATTTGAAGTCCCTTCAAGTTTGAAGTAAAGCATTTTGAAGC 821
DB 1851 ATCGAAATCTACATGGAATTAATTTGAAGTCCCTTCAAGTTTGAAGTAAAGCATTTTGAAGC 1910
```


This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 29, 2004, 05:04:37 ; Search time 4998 Seconds

(without alignments)
11184.874 Million cell updates/sec

Title: US-09-787-491B-18

Perfect score: 1872

Sequence: 1 gccctctagcgtgtgtgtc.....tatcaaaaaaaaaaaaaa 1872

Scoring table: OLIGO_NUC

Gapop 60.0 , Gapext 60.0

Searched: 27513289 seqs, 14931090276 residues

Word size : 0

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: listing first 45 summaries

Database :

EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hic:*
9: gb_est1:*
10: gb_est2:*
11: gb_hic:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_esthum:*
16: em_estom:*
17: em_gss_hum:*
18: em_gss_inv:*
19: em_gss_pln:*
20: em_gss_vrt:*
21: em_gss_fun:*
22: em_gss_man:*
23: em_gss_mus:*
24: em_gss_pro:*
25: em_gss_rtd:*
26: em_gss_phg:*
27: em_gss_vrt:*
28: gb_gss1:*
29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match length	ID	Description
1	794	42.4	914	BX352678 BX352678
2	776	41.5	867	BX392494 BX392494
3	769	41.1	1106	BX363934 BX363934
4	754	40.3	1201	BX356666 BX356666

5	722	38.6	858	9	AU130178	AU130178
6	699	37.3	866	13	BX326041	BX326041
7	690	36.9	926	13	BX450740	BX450740
8	677	36.2	879	13	BQ421218	AGENCOURT
9	657	35.1	1035	12	BM928059	AGENCOURT
10	652	34.8	788	12	BG611131	AGENCOURT
11	642	34.3	710	13	BM683427	UI-CF-RCL
12	640	34.2	1038	12	BM921276	UI-CF-RCL
13	637	34.0	895	13	BX408797	AGENCOURT
14	603	32.2	670	13	BU621746	AGENCOURT
15	596	31.8	760	12	BM728917	UI-E-FL1-
16	595	31.0	761	12	BI836024	UI-E-FL1-
17	580	31.0	679	13	BU742938	603084008
18	573	30.6	727	14	CB053447	UI-E-FL1-
19	560	29.9	622	12	BM836931	NISC_g114
20	560	28.9	871	14	CD512033	AGENCOURT
21	560	29.9	1095	12	BM910977	AGENCOURT
22	557	29.8	786	10	BE281165	AGENCOURT
23	556	29.7	912	13	BQ220979	AGENCOURT
24	553	29.5	579	13	BU689661	UI-CF-RNO
25	551	29.4	1229	12	BM467160	AGENCOURT
26	548	29.3	686	13	BX509477	AGENCOURT
27	539	28.8	673	14	CD690132	EST6655_h
28	539	28.8	722	13	BU928568	AGENCOURT
29	534	28.5	823	12	BI559513	603252842
30	530	28.3	544	12	BM147285	TCAP1010
31	530	28.3	648	12	BM837276	AGENCOURT
32	530	28.3	901	13	BU193345	K-EST0113
33	530	28.3	1114	12	BM544074	AGENCOURT
34	529	28.3	536	10	BE018162	hb76a11.y
35	529	28.3	1068	12	BM456685	AGENCOURT
36	527	28.2	527	14	CB112315	K-EST0154
37	521	27.8	728	10	BE259920	601184810
38	515	27.5	917	13	BX388512	AGENCOURT
39	511	27.3	701	10	BE295413	601175939
40	511	27.3	767	10	BF691855	602248438
41	510	27.2	928	13	BQ428125	AGENCOURT
42	509	27.2	702	9	AU132812	AU132812
43	507	27.1	513	10	BE301644	601184810
44	503	26.9	810	13	BU599405	AGENCOURT
45	499	26.7	699	12	BG532282	602561373

ALIGNMENTS

RESULT 1
LOCUS BX352678
DEFINITION BX352678 Homo sapiens NEUROBLASTOMA COT 10-NORMALIZED Homo sapiens
CDNA clone CSDB001YG15 5-PRIME, mRNA sequence.
ACCESSION BX352678
VERSION BX352678.1 GI:30347694
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens

REFERENCE
AUTHORS Li, W.-B., Gruber, C., Jessee, D., and Polayes, D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished (2001)
COMMENT Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: segret@genoscope.cns.fr, Web: www.genoscope.cns.fr
Library was constructed by Life Technologies, a division of
Invitrogen. This sequence belongs to sequence cluster 6566.r
Contact: Feng Liang Email: fliang@life.com URL:
http://fulllength.invitrogen.com/invitrogen Corporation 1600
Faraday Avenue Genoscope sequence ID: CSDB001AD08QPl.

FEATURES
source
1..914

Db 507 CTAGTGAATCATTACAGTATGGCTCATGTTGTTTGAGTGCTGTTATTCATACAT 448
Qy 1412 ATTTTACACCACTTCGTATCATATGTAATTAAGAACATATATACATCAAGATAAGTA 1471
Db 447 ATTTTACACCACTTCGTATCATATGTAATTAAGAACATATATACATCAAGATAAGTA 388
Qy 1472 TTGTGTGGTATCTGCGCATTTTAAAGTATCCGATTTTGTATCAATTAATTAATATG 1531
Db 387 TTGTGTGGTATCTGCGCATTTTAAAGTATCCGATTTTGTATCAATTAATTAATATG 328
Qy 1532 AAAAATGATTTAATCTGTATATAA CTGGTTATTTGTGAGAGCACTGATTAATTAATG 1591
Db 327 AAAAATGATTTAATCTGTATATAA CTGGTTATTTGTGAGAGCACTGATTAATTAATG 268
Qy 1592 TTATATAAATTTTACTCTGCTCCACCAACACATGCTAGATATAACCCCAAAATA 1651
Db 267 TTATATAAATTTTACTCTGCTCCACCAACACATGCTAGATATAACCCCAAAATA 208
Qy 1652 AGTATTTAATCTGATTAAGTATTAAGAGACCTGGTGCTAATAATTAGATTATTTGAG 1711
Db 207 AGTATTTAATCTGATTAAGTATTAAGAGACCTGGTGCTAATAATTAGATTATTTGAG 148
Qy 1712 GCAGACAGAGAGCTGTATACCTAATCTAATTAATTAATTAATTAATTAATTAATTAAT 1771
Db 147 GCAGACAGAGAGCTGTATACCTAATCTAATTAATTAATTAATTAATTAATTAATTAAT 88
Qy 1772 CCAATTAATCTTTTATGATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1818
Db 87 CCAATTAATCTTTTATGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 41

RESULT 3
EX363934 1106 bp mRNA linear EST 05-MAY-2003
LOCUS EX363934 Homo sapiens B CELLS (RAMOS CELL LINE) COT 25-NORMALIZED
DEFINITION Homo sapiens cDNA clone CS0DL007YC18 5-PRIME, mRNA sequence.
ACCESSION BX363934
VERSION BX363934.1 GI:30378776
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1106)
AUTHORS Li, W.B., Gruber, C., Jessee, J. and Polayes, D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished (2001)
COMMENT Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
Library was constructed by Life Technologies, a division of
Invitrogen. This sequence belongs to sequence cluster 6566.r,
Contact: Feng Liang Email : fliang@life.com URL :
http://fulllength.invitrogen.com/Invitrogen Corporation 1600
Faraday Avenue Genoscope sequence ID : CS0DL007B809P1.
FEATURES
Source
1..1106
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0DL007YC18"
/cell_type="B CELLS (RAMOS CELL LINE) COT 25-NORMALIZED"
/cell_line="RAMOS CELL LINE"
/clone_id="Homo sapiens B CELLS (RAMOS CELL LINE) COT
25-NORMALIZED"
/note="1st strand cDNA was primed with a NotI-oligo (dT)
primer. Five prime end enriched, double-strand cDNA was
digested with Not I and cloned into the Not I and EcoR V
sites of the pCMVSPORT 6 vector. Library was normalized."

ORIGIN
Query Match 41.1%; Score 769; Db 13; Length 1106;

Best Local Similarity 99.8%; Pred. No. 3,1e-170;
Matches 939; Conservative 0; Mismatches 1; Indels 1; Gaps 1;
Qy 70 GCCGCCCGGAGACACGCGCGCCGCTATGCTCCGCTCCCGCAACGCTCTG 129
Db 87 GCCGCCCGGAGACACGCGCGCCGCTATGCTCCGCTCCCGCAACGCTCTG 146
Qy 130 TTGCTCAGGAAAGTGGCGGAGACACGAGTGTGAAGCTTGGGCGGAAATTTGGT 189
Db 147 TTGCTCAGGAAAGTGGCGGAGACACGAGTGTGAAGCTTGGGCGGAAATTTGGT 206
Qy 190 TATGCTCCTAATAGTATGATGTATGCTTCCATTTCTACATCTGCGCTCAAGAGA 249
Db 207 TATGCTCCTAATAGTATGATGTATGCTTCCATTTCTACATCTGCGCTCAAGAGA 266
Qy 250 TTGCTTATGTTCAATTTTGAAGATCTTGTATGCTGAAGCGCTTTAATTAATTGAC 309
Db 267 TTGCTTATGTTCAATTTTGAAGATCTTGTATGCTGAAGCGCTTTAATTAATTGAC 326
Qy 310 AGAAAGTGAATTTGTGACGCGCAATGAATACAGTTTGGCCGAGGGGATCGAAGACA 369
Db 327 AGAAAGTGAATTTGTGACGCGCAATGAATACAGTTTGGCCGAGGGGATCGAAGACA 386
Qy 370 CCAATATGATGAAAGCCCAAGAAAGGAGATGTATACAGTTCTTCAAGCTATGATAT 429
Db 387 CCAATATGATGAAAGCCCAAGAAAGGAGATGTATACAGTTCTTCAAGCTATGATAT 446
Qy 430 TATGACGATACACAGCTTCTTGAAGCCGAAAGTTATGAAGAAGATCAAGAAAGTGG 489
Db 447 TATGACGATACACAGCTTCTTGAAGCCGAAAGTTATGAAGAAGATCAAGAAAGTGG 506
Qy 490 TCTTTGATTTACACTATGAAAGATCGTATAGTCTTGAAGACAGTACGAGCTGGAAGA 549
Db 507 TCTTTGATTTACACTATGAAAGATCGTATAGTCTTGAAGACAGTACGAGCTGGAAGA 566
Qy 550 CCACGGCGTGAAGA-AGCCATTCGACATGATGAACCAATCTGAGTGAATACCCAG 608
Db 567 CCACGGCGTGAAGAAGCCATTCGACATGATGAACCAATCTGAGTGAATACCCAG 626
Qy 609 TACAGTTCTGCTTCTACACTTCAAGAAAGATCTGAAAGCCGAAAGAAACCAAGAAAG 668
Db 627 TACAGTTCTGCTTCTACACTTCAAGAAAGATCTGAAAGCCGAAAGAAACCAAGAAAG 686
Qy 669 GCACTTCAAGGACCAAGAGGTGGTGAAGGTCTGAGATGAATATCTAGAAATAT 728
Db 687 GCACTTCAAGGACCAAGAGGTGGTGAAGGTCTGAGATGAATATCTAGAAATAT 746
Qy 729 TTGACTCTGCTTGAAGAAAGATTAAGATGTTATCGAAACCTACATGATTAATTGAAG 788
Db 747 TTGACTCTGCTTGAAGAAAGATTAAGATGTTATCGAAACCTACATGATTAATTGAAG 806
Qy 789 TCCCTTTAAGTTGAAGATGAAGCTTTTGAAGACAAATTAAGAAATTTCAACTTTGACT 848
Db 807 TCCCTTTAAGTTGAAGATGAAGCTTTTGAAGACAAATTAAGAAATTTCAACTTTGACT 866
Qy 849 TGTGAACTAATATCCCTAATATGATATGATTTATGATTTATGATTTATGATTTATGAT 908
Db 867 TGTGAACTAATATCCCTAATATGATATGATTTATGATTTATGATTTATGATTTATGAT 926
Qy 909 AATTAATTAATTTGAAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 968
Db 927 AATTAATTAATTTGAAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 986
Qy 969 GTGCTCTGTTGTGCTGCTCAACATGAATTTGGTGGGAAA 1009
Db 987 GTGCTCTGTTGTGCTGCTCAACATGAATTTGGTGGGAAA 1027

RESULT 4
EX356666/c 1201 bp mRNA linear EST 05-MAY-2003
LOCUS EX356666 Homo sapiens PLACENTA COT 25-NORMALIZED Homo sapiens cDNA
DEFINITION Clone CS0DI015YD03 3-PRIME, mRNA sequence.

ACCESSION BX356666
 VERSION BX356666.1 GI:30370049
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 1201)
 Li, W.B., Gruber, C., Jessee, J., and Polayes, D.
 Full-length cDNA libraries and normalization
 Unpublished (2001)
 JOURNAL
 COMMENT
 Genoscope - Centre National de Sequencage
 BP 191 91006 Evry cedex - France
 Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
 Library was constructed by Life Technologies, a division of
 Invitrogen. This sequence belongs to sequence cluster 6566.x
 Contact : Feng Liang Email : fliang@lifetech.com URL :
 http://fulllength.invitrogen.com/ Invitrogen Corporation 1600
 Faraday Avenue Genoscope sequence ID : CS0D1015CB02NP1.

FEATURES
 source
 1..1201
 Location/Qualifiers
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="CS0D1015YD03"
 /cissue_type="PLACENTA COT 25-NORMALIZED"
 /clone_lib="Homo sapiens PLACENTA COT 25-NORMALIZED"
 /note="1st strand cDNA was primed with a NotI-oligo(dT)
 primer. Five prime end enriched, double-strand cDNA was
 digested with Not I and cloned into the Not I and EcoR V
 sites of the pCMVSPORT 6 vector. Library was normalized."

ORIGIN

Query Match 40.3%; Score 754; DB 13; Length 1201;
 Best Local Similarity 99.7%; Pred. No. 9e-167;
 Matches 974; Conservative 0; Mismatches 2; Indels 1; Gaps 1;

853 GAACTAATCCCTAATATGAATAGTTATATGATTCATGGGTACAGGTCCTAATA 912
 993 GAACTAATCCCTAATATGAATAGTTATATGATTCATGGGTACAGGTCCTAATA 934
 913 AATTATGGAACCTGGATGCTGAATATCAAGGAAGACAGCTGCTCTTACAGTGC 972
 933 AATTATGGAACCTGGATGCTGAATATCAAGGAAGACAGCTGCTCTTACAGTGC 874
 973 CTCTGTGTGCTGTCTCAACTGAATGGGTGGGAAAGATGTGCTCAATATTAAGTT 1032
 873 CTCTGTGTGCTGTCTCAACTGAATGGGTGGGAAAGATGTGCTCAATATTAAGTT 814
 1033 CCATTTTGGCATTATGGCAATCTTGGCTTTGTTATTTGGTCCAGTGTCTTCTGC 1092
 813 CCATTTTGGCATTATGGCAATCTTGGCTTTGTTATTTGGTCCAGTGTCTTCTGC 754
 1093 TTAATCATTTGGCTTTGTGGCACTGTGTTATTTACTGTGACACACAGCAGTTTAA 1152
 753 TTAATCATTTGGCTTTGTGGCACTGTGTTATTTACTGTGACACACAGCAGTTTAA 694
 1153 TCTGTCTTAATACTCTCTCCAGTAAATTCATATATTTGACATCCAGCTAAGAG 1212
 693 TCTGTCTTAATACTCTCTCCAGTAAATTCATATATTTGACATCCAGCTAAGAG 634
 1213 GCCCATCTCTTGCACACTCTTCTCTAGTCAATATTTGCAAAATTTTATGAGCCCT 1272
 633 GCCCATCTCTTGCACACTCTTCTCTAGTCAATATTTGCAAAATTTTATGAGCCCT 574
 1273 ACTGTGGCAAAATCATTTGATCTGATATTTGCAAAATATAGTAATTCCTTATTCAGTA 1332
 573 ACTGTGGCAAAATCATTTGATCTGATATTTGCAAAATATAGTAATTCCTTATTCAGTA 514
 1333 AA-TGTCTACTGACCAATCTAGTAATCATTTACAGTATGGCTCATTTGTTTGA 1391
 513 AATTGTATCTGACCAATCTAGTAATCATTTACAGTATGGCTCATTTGTTTGA 454

1392 GGTGTATTTCATACATATTTTACCACTGCTATCATATGTAATTTAGAACAT 1451
 453 GGTGTATTTCATACATATTTTACCACTGCTATCATATGTAATTTAGAACAT 394
 1452 ATACATCAAGGATTAAGTAATTTGTGGTAACTGCTCAATTTAAAGTATCCAGTATTGA 1511
 393 ATACATCAAGGATTAAGTAATTTGTGGTAACTGCTCAATTTAAAGTATCCAGTATTGA 334
 1512 TCACATTTATTAATTAATGAAAAATGATTTAATCTGATATTAACCTGTTATTGTGCA 1571
 333 TCACATTTATTAATTAATGAAAAATGATTTAATCTGATATTAACCTGTTATTGTGCA 274
 1572 GTGACTGTAATTAATTAAGTATTAATTAATTTGTTACCTGCTCCACCAACACATGCT 1631
 273 GTGACTGTAATTAATTAAGTATTAATTAATTTGTTACCTGCTCCACCAACACATGCT 214
 1632 AGGATATTAACCCCAAAATTAAGTATTTAATCTTGAATTAAGGAGAGCTGGTGC 1691
 213 AGGATATTAACCCCAAAATTAAGTATTTAATCTTGAATTAAGGAGAGCTGGTGC 154
 1692 TATTAATTAATTTATTTGAGGACAGACAGAGCTGTATCTTAATGATTTAGTATGTTTC 1751
 153 TATTAATTAATTTATTTGAGGACAGACAGAGCTGTATCTTAATGATTTAGTATGTTTC 94
 1752 TGTATTTGAAAAAAGTTCACCAATTAATTAATTTTATGATTTACATGATTAATTTATA 1811
 93 TGTATTTGAAAAAAGTTCACCAATTAATTAATTTTATGATTTACATGATTAATTTATA 34
 1812 GGGGACATGTTCTGTGT 1828
 33 GGGGACATGTTCTGTGT 17

RESULT 5
 AUI30178 858 bp mRNA linear EST 01-AUG-2002
 LOCUS AUI30178 NT2RP3 Homo sapiens cDNA clone NT2RP3000370 5', mRNA
 DEFINITION

ACCESSION AUI30178
 VERSION AUI30178.1 GI:10990532
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 858)
 Ota, T., Nishikawa, T., Suzuki, Y., Ishii, S., Saito, K., Kawai, Y.,
 Yamamoto, T., Wakamatsu, A., Nakamura, Y., Nagai, T., Sugano, S. and
 Isogai, T.
 HRI human cDNA project
 Unpublished (2000)
 CONTACT: Takao Isogai
 Genomics Laboratory
 Helix Research Institute
 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
 Tel: 81-438-52-3975
 Fax: 81-438-52-3986
 Email: genomics@hri.co.jp
 HRI human cDNA project; 5'- & 3'-end one pass sequencing; Helix
 Research Institute; cDNA library construction; Department of
 Virology, Institute of Medical Science, University of Tokyo, and
 Helix Research Institute.

FEATURES
 source
 1..858
 Location/Qualifiers
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="NT2RP3000370"
 /cell_type="teratocarcinoma"
 /clone_lib="NT2RP3"
 /note="Vector: PME18SFL3; mRNA from NT2 neuronal precursor"

1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
 Tel: 81-438-52-3975
 Fax: 81-438-52-3986
 Email: genomics@hri.co.jp
 HRI human cDNA project; 5'- & 3'-end one pass sequencing; Helix
 Research Institute; cDNA library construction; Department of
 Virology, Institute of Medical Science, University of Tokyo, and
 Helix Research Institute.

ORIGIN cells after 2-weeks retinoic acid (RA) induction"

Query Match 38.6%; Score 722; DB 9; Length 858;
Best Local Similarity 100.0%; Pred. No. 3,4e-159; Indels 0; Gaps 0;
Matches 722; Conservative 0; Mismatches 0;

QY 1080 CAGTGTTCCTGCTTAATCATTTGCTTGTGTGGCATCTGTGTTATTTACTGTACACCA 1139
DB 1 CAGTGTTCCTGCTTAATCATTTGCTTGTGTGGCATCTGTGTTATTTACTGTACACCA 60
QY 1140 CAGGAGTTTACATCTGTCTTAACCTCTCTCCGAGTAAATTCGAATTAATTGACA 1199
DB 61 CAGGAGTTTACATCTGTCTTAACCTCTCTCCGAGTAAATTCGAATTAATTGACA 120
QY 1200 TCCAGCTAAGAGGGCCCATCTCTTCCACCTCTTCCAGTATATTTAGCAAAAT 1259
DB 121 TCCAGCTAAGAGGGCCCATCTCTTCCACCTCTTCCAGTATATTTAGCAAAAT 180
QY 1260 TTATTGAGCCCTTACTGTGTGGCAAAATCATTTGATGATTAATTGAAAAATGATAAT 1319
DB 181 TTATTGAGCCCTTACTGTGTGGCAAAATCATTTGATGATTAATTGAAAAATGATAAT 240
QY 1320 CCTTATTCAGTAATNGCTACTGTAGCAGCAATCTAGTATCATTTACAGTATGCTCAT 1379
DB 241 CCTTATTCAGTAATNGCTACTGTAGCAGCAATCTAGTATCATTTACAGTATGCTCAT 300
QY 1380 TGTGTTGTTGGAGTGTGTTATTCATTAACAATATTTTACACCATTCGATCAATGTAAT 1439
DB 301 TGTGTTGTTGGAGTGTGTTATTCATTAACAATATTTTACACCATTCGATCAATGTAAT 360
QY 1440 ATGAGCAACATATATCATATCAAGATTAAGTATGTGTGTTATTCGATTAATTAAGTA 1499
DB 361 ATGAGCAACATATATCATATCAAGATTAAGTATGTGTGTTATTCGATTAATTAAGTA 420
QY 1500 TCCAGATATTTGATCATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1559
DB 421 TCCAGATATTTGATCATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 480
QY 1560 GTTTATTTGTCAGTACTGTATATATCTAGAGTTTAAATTTTACTCTGCTCAG 1619
DB 481 GTTTATTTGTCAGTACTGTATATATCTAGAGTTTAAATTTTACTCTGCTCAG 540
QY 1620 CAACACATGCTAGGATATTAACCCCAAAATTAAGTATTAATTTGATTAATTAAG 1679
DB 541 CAACACATGCTAGGATATTAACCCCAAAATTAAGTATTAATTTGATTAATTAAG 600
QY 1680 GAGACTGGGTGCTATTAATTAATTTTGAAGCAGACAGAGCTGTTATCTTAACCTGA 1739
DB 601 GAGACTGGGTGCTATTAATTAATTTTGAAGCAGACAGAGCTGTTATCTTAACCTGA 660
QY 1740 TTATGATATGTTCTGTAATTTGAAAAATGTCAACAAATTAATTTTATGATTAAT 1799
DB 661 TTATGATATGTTCTGTAATTTGAAAAATGTCAACAAATTAATTTTATGATTAAT 720
QY 1800 GT 1801
DB 721 GT 722

RESULT 6
EX326041/c 866 bp mRNA linear EST 02-MAY-2003
LOCUS EX326041 Homo sapiens NEUROBLASTOMA COR 10-NORMALIZED Homo sapiens
DEFINITION CDNA clone CS0DB001YGL5 3-PRIME, mRNA sequence.
ACCESSION EX326041
VERSION EX326041.1 GI:30344506
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 866)

AUTHORS Li, M.B., Gruber, C., Jesse, J. and Polayes, D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished (2001)
COMMENT Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: seqref@genoscope.cns.fr Web : www.genoscope.cns.fr
Library was constructed by life technologies, a division of
Invitrogen. This sequence belongs to sequence cluster 6366.r
Contact : Feng Liang Email : fliang@life.com URL :
http://fulllength.invitrogen.com/Invitrogen Corporation 1600
Faraday Avenue Genoscope sequence ID : CS0A0022C11_U0179_1.
Location/Qualifiers
1..866
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0DB001YGL5"
/tissue.type="NEUROBLASTOMA COR 10-NORMALIZED"
/clone.lib="Homo sapiens NEUROBLASTOMA COR 10-NORMALIZED"
/note="First strand cDNA was primed with a NotI-oligo(dT)
primer. Five prime end enriched, double-strand cDNA was
digested with Not I and cloned into the Not I and EcoR V
sites of the pCMVSPORT 6 vector. Library was normalized."

FEATURES
source

ORIGIN

Query Match 37.3%; Score 699; DB 13; Length 866;
Best Local Similarity 99.9%; Pred. No. 7.6e-154; Indels 0; Gaps 0;
Matches 749; Conservative 0; Mismatches 1;

QY 561 AGAAGCCATTCGCAATGATATGACCAAACTGCACTGGAATACCGATACAGTCTGCT 620
DB 788 AGAAGCCATTCGCAATGATATGACCAAACTGCACTGGAATACCGATACAGTCTGCT 729
QY 621 TACTACCTTCAAGAAAGATCTGAAAGCGGAAAAAGAAACCAAGAGGCGAGTTCAGCG 680
DB 728 TACTACCTTCAAGAAAGATCTGAAAGCGGAAAAAGAAACCAAGAGGCGAGTTCAGCG 669
QY 681 ACCAAGGGGTGGGAGGAGGTCGCACTATGAATCTGTACGAATATTTGACTCTGCT 740
DB 668 ACCAAGGGGTGGGAGGAGGTCGCACTATGAATCTGTACGAATATTTGACTCTGCT 609
QY 741 CTGAAAAGATTAAGAAATGTTATCGAAATCATCATGATTAATTAATTAATTAATTAAT 800
DB 608 CTGAAAAGATTAAGAAATGTTATCGAAATCATCATGATTAATTAATTAATTAATTAAT 549
QY 801 TGAAGTATGATTTAGACCAAAATTAAGAAATTCACCTTGTACTTGTGAAACCTGA 860
DB 548 TGAAGTATGATTTAGACCAAAATTAAGAAATTCACCTTGTACTTGTGAAACCTGA 489
QY 861 TCCCTAATATGAATAGGTTTATATGATTCATGAGTAAACAGTCCATTAATTAATTAAT 920
DB 488 TCCCTAATATGAATAGGTTTATATGATTCATGAGTAAACAGTCCATTAATTAATTAAT 429
QY 921 GAAACTAGATGTCTGAATATCAAGAAAGACAGCATAGTCTTTACAGTCCCTGCTTG 980
DB 428 GAAACTAGATGTCTGAATATCAAGAAAGACAGCATAGTCTTTACAGTCCCTGCTTG 369
QY 981 GTCTGTCTCAAACTGAATTTGGGTGGGAAAGGTAATGTCGAATTAATTAATTAATTT 1040
DB 368 GTCTGTCTCAAACTGAATTTGGGTGGGAAAGGTAATGTCGAATTAATTAATTAATTT 309
QY 1041 GCAATTAATGAGCAATCTGCTTGTATTTTGTGCGGAGTCTTCTGCTTAATCAT 1100
DB 308 GCAATTAATGAGCAATCTGCTTGTATTTTGTGCGGAGTCTTCTGCTTAATCAT 249
QY 1101 TTGCTTTGCGCATCTGTGTTTATTTACTTTGACACCAATGAGTTCATCTGCTT 1160
DB 248 TTGCTTTGCGCATCTGTGTTTATTTACTTTGACACCAATGAGTTCATCTGCTT 189
QY 1161 AACTACTCTTCCAGGTAATTCGAATTAATTTGACATTCAGCTAAGAGGCGCCATCT 1220
DB 188 AACTACTCTTCCAGGTAATTCGAATTAATTTGACATTCAGCTAAGAGGCGCCATCT 129

QY 1221 CTTCTGACCTCTTCTTCTAGTCGATATTCAGCAATATTTATGAGCCCTTACTGTGGG 1280
DB 128 CTTCTGACCTCTTCTTCTAGTCGATATTCAGCAATATTTATGAGCCCTTACTGTGGG 69
QY 1281 CAAATCATGTACTGATATGAGAAAA 1310
DB 68 CAAATCATGTACTGATATGAGAAAA 39

RESULT 7
BX450740 926 bp mRNA linear EST 22-MAY-2003
LOCUS BX450740 Homo sapiens B CELLS (RAMOS CELL LINE) Homo sapiens cDNA
DEFINITION clone CS0D501YH09 5-PRIME, mRNA sequence.
ACCESSION BX450740 GI:31022152
VERSION BX450740.1
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (bases 1 to 926)
AUTHORS Li, W.B., Gruber, C., Jesse, J., and Polayes, D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished (2001)
COMMENT Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: sequef@genoscope.cns.fr, Web: www.genoscope.cns.fr
Library was constructed by Life Technologies, a division of
Invitrogen. This sequence belongs to sequence cluster 798.r for
more information about this cluster, see
http://www.genoscope.cns.fr/
cgi-bin/cluster.cgi?seq=CS0BA011ZH09_CS01033_1cluster=798.r.
Contact: Feng Liang Email: fliang@life.techn.com URL:
http://fulllength.invitrogen.com/ Invitrogen Corporation 1600
Faraday Avenue Genoscope sequence ID: CS0BA011ZH09_CS01033_1.

FEATURES
source
1..926
Location/Qualifiers
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0D501YH09"
/tissue_type="B CELLS (RAMOS CELL LINE)"
/cell_line="RAMOS CELL LINE"
/clone_1lb="Homo sapiens B CELLS (RAMOS CELL LINE)"
/note="Vector: pCMVSPORT 6; 1st strand cDNA was primed
with a NotI-oligo(dT) primer. Five prime end enriched,
double-strand cDNA was digested with Not I and cloned into
the Not I and EcoRV sites of the pCMVSPORT 6 vector.
Library was not normalized."

ORIGIN
Query Match 36.9%; Score 690; DB 13; Length 926;
Best Local Similarity 100.0%; Pred. No. 8.9e-152;
Matches 690; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 779 AATAATGAGTCCCTTCAAGTTTGAAGTAAAGTAAATTTAGACAAATTAAGAAATTTCA 838
DB 21 AATAATGAGTCCCTTCAAGTTTGAAGTAAAGTAAATTTAGACAAATTAAGAAATTTCA 80
QY 839 ACTTTGACTGTGGGAACTTAATCCCTTAATATGAAATAGTTATTTGATTCATGGGTA 898
DB 81 ACTTTGACTGTGGGAACTTAATCCCTTAATATGAAATAGTTATTTGATTCATGGGTA 140
QY 899 ACAGTCCATATAAATTAATTTGAAACTAGATGTGATATCAAGAGACAGCCATA 958
DB 141 ACAGTCCATATAAATTAATTTGAAACTAGATGTGATATCAAGAGACAGCCATA 200
QY 959 GTCTCTTACAGTCCCTGTGTGTCTGTCTCAAACTGAATTTGGTGGGAAAAAGGTATGT 1018
DB 201 GTCTCTTACAGTCCCTGTGTGTGTCTGTCTCAAACTGAATTTGGTGGGAAAAAGGTATGT 260

QY 1019 CCAATATAAANGTCCATTTTGGCATATTTGGCAAAATCTGCTTTTATTTTGGTG 1078
DB 261 CCAATATAAANGTCCATTTTGGCATATTTGGCAAAATCTGCTTTTATTTTGGTG 320
QY 1079 CCAAGTCTTCTGCTTAAATTCATTTGCTTTGTTGGCATCTGTGTTTATTTACTTGGAC 1138
DB 321 CCAAGTCTTCTGCTTAAATTCATTTGCTTTGTTGGCATCTGTGTTTATTTACTTGGAC 380
QY 1139 ACATGCAAGTTTACATCTGTCTTAACTACTCTTTCCAGGTAAATTCATTTATTTGAC 1198
DB 381 ACATGCAAGTTTACATCTGTCTTAACTACTCTTTCCAGGTAAATTCATTTATTTGAC 440
QY 1199 ATCCAGCTTAAGAGGCGCCATCTCTCTGACCTCTTTCCAGTAAATTTAGCAAAAT 1258
DB 441 ATCCAGCTTAAGAGGCGCCATCTCTCTGACCTCTTTCCAGTAAATTTAGCAAAAT 500
QY 1259 TTTATTTAGGCCCTTACTGTGGCAATCATTTGTACTGTATTTGAAGAAATATGATAAT 1318
DB 501 TTTATTTAGGCCCTTACTGTGGCAATCATTTGTACTGTATTTGAAGAAATATGATAAT 560
QY 1319 TCCCTTATTCAGTAAATGCTCTCTGAGCAATCTTATGTAATCATTTACGTATGGCTCA 1378
DB 561 TCCCTTATTCAGTAAATGCTCTCTGAGCAATCTTATGTAATCATTTACGTATGGCTCA 620
QY 1379 TTGTTTGTGTTGAGGTGTGTATTTCAATAAATTTTACACATTCGTATCAATGTAA 1438
DB 621 TTGTTTGTGTTGAGGTGTGTATTTCAATAAATTTTACACATTCGTATCAATGTAA 680
QY 1439 TATAGAACAAATATATCATCATCAAGATAG 1468
DB 681 TATAGAACAAATATATCATCATCAAGATAG 710

RESULT 8
BQ421218 879 bp mRNA linear EST 23-MAY-2002
LOCUS BQ421218
DEFINITION AGENCOURT 7763286 NIH_MGC_92 Homo sapiens cDNA clone IMAGE:6013472
5', mRNA sequence.
ACCESSION BQ421218
KEYWORD BQ421218.1 GI:21116533
VERSION EST.
KEYWORDS BQ421218.1 GI:21116533
SOURCE Homo sapiens
ORGANISM Homo sapiens (human)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (bases 1 to 879)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/
TITLE NIH-MGC National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cga@bbs-r@mail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LINL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LINL at:
http://image.llnl.gov
Plate: LLAM13206 row: g column: 09
High quality sequence stop: 679.
Location/Qualifiers
1..879
Location/Qualifiers
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:6013472"
/tissue_type="embryonal carcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/clone_1lb="NIH MGC 92"
/note="Organ: testis; Vector: pCMV-SPORT6; Site 1: NotI;
Site 2: SalI; Cloned unidirectionally; Oligo-dT primed.
Average insert size 2.5 kb. Library enriched for

ORIGIN

full-length clones and constructed by Life Technologies.
Note: this is a NIH_MGC library."

Query Match 36.2%; Score 677; DB 13; Length 879;
Best Local Similarity 99.7%; Pred. No. 9,9e-149;
Matches 777; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```
QY 932 GTCTGAATATCAAGGAAGACAGCCATAGTCTCTTACAGTGCCTCTGTGTCTCAA 991
DB 5 GTCTGAATATCAAGGAAGACAGCCATAGTCTCTTACAGTGCCTCTGTGTCTCAA 64
QY 992 ACTGAATGGGGGGGAAAAGGTATGCTCCAAATTAAGTTCATTTTGGCCATTATG 1051
DB 65 ACTGAATGGGGGGGAAAAGGTATGCTCCAAATTAAGTTCATTTTGGCCATTATG 124
QY 1052 CAAATCTTGCCCTTTGTTATTTTGGGCGCAGTGTCTTGGCTTAACATTTGTTG 1111
DB 125 CAAATCTTGCCCTTTGTTATTTTGGGCGCAGTGTCTTGGCTTAACATTTGTTG 184
QY 1112 GCATCTGTGTTATTTACTTGTACACGACAGATTTTACATCTGTCTTACTACTCT 1171
DB 185 GCATCTGTGTTATTTACTTGTACACGACAGATTTTACATCTGTCTTACTACTCT 244
QY 1172 CCGAGGTAATCCCAATTTATTTGACATCCGCTTAAGGGGCCATCTCTTCTCACTC 1231
DB 245 CCGAGGTAATCCCAATTTATTTGACATCCGCTTAAGGGGCCATCTCTTCTCACTC 304
QY 1232 TTTCCTAGTCAATATTTGACGCAATATTTATTTGAGCCCTTACTGTGGGCAATATG 1291
DB 305 TTTCCTAGTCAATATTTGACGCAATATTTATTTGAGCCCTTACTGTGGGCAATATG 364
QY 1292 ACTGATTAATGAGAAAATAGATTAATTCCTTATTCAGTAATGTCTACTGAGCAAT 1351
DB 365 ACTGATTAATGAGAAAATAGATTAATTCCTTATTCAGTAATGTCTACTGAGCAAT 424
QY 1352 CTAAGGATCATTAAGTATGAGTGGCCCATTTGTTGTTGAGGTGTTATCATTAACAT 1411
DB 425 CTAAGGATCATTAAGTATGAGTGGCCCATTTGTTGAGGTGTTATCATTAACAT 484
QY 1412 ATTTTACACATTCCTATCAATGTATTTATAGAACACAAATATAGATCAAGGATAAGTA 1471
DB 485 ATTTTACACATTCCTATCAATGTATTTATAGAACACAAATATAGATCAAGGATAAGTA 544
QY 1472 TTGTGTGTTATCTGCCATTTAAAGTATCCAGTATTTGATCACAATTATTAATAATG 1531
DB 545 TTGTGTGTTATCTGCCATTTAAAGTATCCAGTATTTGATCACAATTATTAATAATG 604
QY 1532 AAAAATGATTTAATCTGTATTAATTAATCTGTTATTTGTCAGTGTATTAATTAAG 1591
DB 605 AAAAATGATTTAATCTGTATTAATTAATCTGTTATTTGTCAGTGTATTAATTAAG 664
QY 1592 TTATATTAATTTGTTTACTCTGCCACCAACACATCTAGGATATTAACCCCAAAATA 1651
DB 665 TTATATTAATTTGTTTACTCTGCCACCAACACATCTAGGATATTAACCCCAAAATA 724
QY 1652 AGTATTTACTTTGATTAAGGTATTAAGAGACTGGGTCTATTAATTAGATTATTTGA 1710
DB 725 GGTATTTAACTTGCATTAGGATTAAGAGACTGGGTCTATTAATTAGATTATTTGA 783
```

RESULT 9

BM928059

LOCUS BM928059 1035 bp mRNA linear EST 12-MAR-2002

DEFINITION AGENCOURT 6728747 NIH_MGC_100 Homo sapiens cDNA clone IMAGE:5797875

5', mRNA sequence.

ACCESSION

BM928059

VERSION BM928059.1 GI:19378438

KEYWORDS

SOURCE EST. Homo sapiens (human)

ORGANISM

Homo sapiens (human)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 1035)
AUTHORS NIH-MGC
TITLE NIH-MGC http://img.ncbi.nih.gov/
JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)
COMMENT Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgaabs-remail.nih.gov
Tissue Procurement: CGAP (Stanford)
CDNA Library Preparation: Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LINL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LINL at:
http://image.llnl.gov
Plate: LICM2024 row: p column: 04
High quality sequence scop: 678.

FEATURES

source

1..1035
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:5797875"
/tissue_type="hepatocellular carcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH MGC 100"
/note="Organ: liver; Vector: pOTB7; Site 1: XhoI; Site 2:
EcoRI; cDNA made by oligo-dT priming. Directionally cloned
into EcoRI/XhoI sites using the following 5' adaptor:
GGCAGAC(G). Size-selected >500bp for average insert size
1.8kb. Library constructed by Ling Hong in the laboratory
of Gerald M. Rubin (University of California, Berkeley)
using ZAP-CDNA synthesis kit (Stratagene) and Superscript
II RT (Life Technologies). Note: this is a NIH_MGC
Library."

ORIGIN

Query Match 35.1%; Score 657; DB 12; Length 1035;
Best Local Similarity 99.9%; Pred. No. 3,9e-144;
Matches 777; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

```
QY 215 TTCCACTGATTTTCTACACTGCGGCTCCAAAGAGATTGCTTATGTTCAATTGAGATG 274
DB 1 TTCCACTGATTTTCTACACTGCGGCTCCAAAGAGATTGCTTATGTTCAATTGAGATG 60
QY 275 TTCTGATGCTGAAGAGCGCTTTACATATTTGACAGAAATGTGATTTGTGACGGCAGA 334
DB 61 TTCTGATGCTGAAGAGCGCTTTACATATTTGACAGAAATGTGATTTGTGACGGCAGA 120
QY 335 TTGAATATCAAGTTTCCCGAGGGGATCGAAAGACCAAAATCAATGAAAGCCAAAGAG 394
DB 121 TTGAATATCAAGTTTCCCGAGGGGATCGAAAGACCAAAATCAATGAAAGCCAAAGAG 180
QY 395 GGAGGAATGTGTACAGTTCTTCAAGCTATGTATATATGACAGATACAGAGTTCTAGAA 454
DB 181 GGAGGAATGTGTACAGTTCTTCAAGCTATGTATATATGACAGATACAGAGTTCTAGAA 240
QY 455 GCCAAGTTATGAAGGAGGAGATCAAGAGTGGGTCTTTGATTTACACTATGAAGAT 514
DB 241 GCCAAGTTATGAAGGAGGAGATCAAGAGTGGGTCTTTGATTTACACTATGAAGAT 300
QY 515 GGTATAGTCTTGAAGACATGAGCCGACTGGAAGACCAAGCGCTAG-AGAAGCCATTCCG 573
DB 301 GGTATAGTCTTGAAGACATGAGCCGACTGGAAGACCAAGCGCTAGAGCAAGATTCG 360
QY 574 ACATATATACCAAACTGACGTGGAATATCCCAATCAAGTTCTGCTTACTACACTTCAA 633
DB 361 ACATATATACCAAACTGACGTGGAATATCCCAATCAAGTTCTGCTTACTACACTTCAA 420
QY 634 GAAAGATCTGAAGCGGAAAAAGAACCAAGAGGCGAGTTCAAGCGAACCAAGGAGGG 693
DB 421 GAAAGATCTGAAGCGGAAAAAGAACCAAGAGGCGAGTTCAAGCGAACCAAGGAGGG 480
QY 694 TGGAAGTGTGACGATGTGATATCTGTACGAATATTTGACTGTGCTGAAAAAGATAAA 753
```


Db	481	TGGAAGGTGCTGCAGATATGAATCTGACGAATATTTTGACTCTGCTGTAAGAAAAAGATAAA	540
QY	754	AGAAATGTTATCGAAAACTACATGGAATAAATTGAAGTCCCTTCAGTTTGGAAAGTAAGCAT	813
Db	541	AGAAATGTTATTCGAAAACTACATGGAATAAATTGAAGTCCCTTCAGTTTGGAAAGTAAGCAT	600
QY	814	TTTAGACAAATATAAAGAAATTTCAACTTTGTACTTGTGGAAACTAATCCCTAAATATGA	873
Db	601	TTTAGGACAAATATAAAGAAATTTCAACTTTGTACTTGTGGAAACTAATCCCTAAATATGA	660
QY	874	ATAGGTTTAAATGATTTCATGGGTAACAAGTCCATAATTAATTTTGGAAACTTGGATGT	933
Db	661	ATAGGTTTAAATGATTTCATGGGTAACAAGTCCATAATTAATTTTGGAAACTTGGATGT	720
QY	934	CTGAATATTCAAAGGAAGACAGCCATAGTCTCTTAAGAGCCCTGTGTGGTCTGTCTCAA	991
Db	721	CTGAATATTCAAAGGAAGACAGCCATAGTCTCTTAAGAGTCTGTGTGGTCTGTCTCAA	778

RESULT 10

LOCUS	BG611131
DEFINITION	BG611131 788 bp mRNA linear EST 18-APR-2001
ACCESSION	60261208SF01 NIH_MGC_60 Homo sapiens cDNA clone IMAGE:473735 5',
VERSION	BG611131
KEYWORDS	BG611131.1 GI:13662502
SOURCE	EST.
ORGANISM	Homo sapiens (human)
REFERENCE	Homo sapiens
AUTHORS	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
TITLE	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo..
JOURNAL	1 (bases 1 to 788)
COMMENT	NIH-MGC http://mgc.nci.nih.gov/. National Institutes of Health, Mammalian Gene Collection (MGC) Unpublished (1999) Contact: Robert Strausberg, Ph.D. Email: cgsabrs@email.nih.gov Tissue Procurement: DCTD/DP cDNA Library Preparation: CLONETECH Laboratories, Inc. cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL) DNA Sequencing by: Incyte Genomics, Inc. Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov Plate: ILM1605 row: b column: 24 High quality sequence stop: 679. Location/Qualifiers 1..788

FEATURES

source

ORIGIN				
Query Match	34.8%	Score 652;	DB 12;	Length 788;
Best Local Similarity	99.9%	Pred. No. 7.2e-143;		
Matches 702; Conservative	0;	Mismatches 1;	Indels 0;	Gaps 0;

QY	671	AGTTCAAGGACCAAAAGGGGCGGTGGAAAGGTCGCGAGTGTGAATCTGACGAATATTT	730
Db	1	AGTTCAAGGACCAAAAGGGTGGAAAGGTCGCGAGTGTGAATCTGACGAATATTT	60
QY	731	TGACTCTGCTGTGAAAAGATPAAAAGATGTTATCGAAAACCTACATGGAATTAATTGAAGTC	790
Db	61	TGACTCTGCTGTGAAAAGATPAAAAGATGTTATCGAAAACCTACATGGAATTAATTGAAGTC	120
QY	791	CCCTCAAGGTTGGAAGPAGCATTTTATAGACAATAAAGAAATTCACCTTTGTACTTG	850
Db	121	CCCTCAAGGTTGGAAGPAGCATTTTATAGACAATAAAGAAATTCACCTTTGTACTTG	180
QY	851	TGGAACCTAATCCCTTAATAATGAATAGGTTTAAATTGATTCATGGGTAACAGGTCACATA	910
Db	181	TGGAACCTAATCCCTTAATAATGAATAGGTTTAAATTGATTCATGGGTAACAGGTCACATA	240
QY	911	TAAATTATTTGGAAACTGAGAGTCTGAAATTCAGAGAAAGACGACATAGTCTCTACAGT	970
Db	241	TAAATTATTTGGAAACTGAGAGTCTGAAATTCAGAGAAAGACGACATAGTCTCTACAGT	300
QY	971	GCCCTGTGTGCTGTCTCAAACTGAATTTGGGTGGAAAAAGTATGCTCCCAATATAAAG	1030
Db	301	GCCCTGTGTGCTGTCTCAAACTGAATTTGGGTGGAAAAAGTATGCTCCCAATATAAAG	360
QY	1031	TTCCATTTTGGCCATPAPTTGGCCAACTCTGGCTCTGTTATTTGGGCGCCAGTCTTCT	1090
Db	361	TTCCATTTTGGCCATPAPTTGGCCAACTCTGGCTCTGTTATTTGGGCGCCAGTCTTCT	420
QY	1091	GCTTAATCATTTGGTTTGTGGCACTGTCTGTTATTTTACCTGTACACCAACATGCAGTTTA	1150
Db	421	GCTTAATCATTTGGTTTGTGGCACTGTCTGTTATTTTACCTGTACACCAACATGCAGTTTA	480
QY	1151	CATCTGTCTTACACTCTCTCCACAGTAATTTCCATTAATTAATTGACATCCAGCTAGA	1210
Db	481	CATCTGTCTTACACTCTCTCCACAGTAATTTCCATTAATTAATTGACATCCAGCTAGA	540
QY	1211	GGGGCCCATCTCTTCTCACCTCTTCTCCAGTACAGTATATTCAGCAATATTTATTTAGGCC	1270
Db	541	GGGGCCCATCTCTCTCACCTCTTCTCCAGTACAGTATATTCAGCAATATTTATTTAGGCC	600
QY	1271	TTACTGTGGGCAAAATCATTTGTACTGGATTAATTGAGAAAAATAGATTAATTCCTTATTCAG	1330
Db	601	TTACTGTGGGCAAAATCATTTGTACTGGATTAATTGAGAAAAATAGATTAATTCCTTATTCAG	660
QY	1331	TAAATGTCTACGAGCAACATCTAGTGAATCATTAACAGTATGCG 1373	
Db	661	TAAATGTCTACGAGCAACATCTAGTGAATCATTAACAGTATGCG 703	

REFERENCE	1 (bases 1 to 710)
AUTHORS	Ronaldo,M.F., Lennon,G. and Soares,M.B.
TITLE	Normalization and subtraction: two approaches to facilitate gene
JOURNAL	discovery
MEDLINE	Genome Res. 6 (9), 791-806 (1996)
PUBMED	97044477
COMMENT	8889548
	Contact: McCray, PB
	McCray Lab
	University of Iowa
	2024 University of Iowa Med Labs, Iowa City, IA 52242, USA
	Tel.: 319 356 4866

Fax: 319 356 7171
 Email: paul-mccray@uiowa.edu
 Tissue Procurement: Dr. M. J. Welsh, University of Iowa
 CDNA Library Preparation: Dr. M. Bento Soares, University of Iowa
 CDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
 DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
 Clone Distribution: Researchers may obtain clones from Research
 Genetics (www.reagen.com) or from Open Biosystems
 (www.openbiosystems.com).
 The following repetitive elements were found in this CDNA
 sequence: 579-628, >LINE2
 Seq primer: M13 FORWARD
 POLYA=Yes.

FEATURES

Location/Qualifiers
 1..710
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="UI-CF-EC1-ace-b-14-0-UI"
 /tissue_type="lung"
 /dev_stage="Adult and Fetal"
 /lab_host="DH10B (Life Technologies) (T1 phage resistant)"
 /clone_11b="UI-CF-EC1"
 /note="Organ: Lung; Vector: pT73-Pac (Pharmacia) with a
 modified polylinker; Site 1: EcoR I; Site 2: Not I.
 UI-CF-EC1 is a normalized cDNA library containing the
 following tissue(s): Normal lung from adult and from fetal
 day 64, day 87, week 19 and week 42. The library was
 constructed according to Bonaldo, Lennon and Soares,
 Genome Research, 6:791-806, 1996. First strand cDNA
 synthesis was primed with an oligo-dT primer containing a
 Not I site. Double stranded cDNA was ligated to an EcoR I
 adaptor, digested with Not I, and cloned directionally
 into pT73-Pac vector. The oligonucleotide used to prime
 the synthesis of first-strand cDNA contains a library tag
 sequence that is located between the Not I site and the
 (dT)18 tail. The sequence tag for this library is
 AAGTCTTAC.
 TAG TISSUE=Normal Lung Epithelial Cells Tissue nos 369-371
 and 380-383
 TAG_LIB=UI-CF-EC1
 TAG_SEQ=AAGTCTTAC"

ORIGIN

Query Match 34.3%; Score 642; DB 13; Length 710;
 Best Local Similarity 99.9%; Pred. No. 1,7e-140;
 Matches 692; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

1165 ACTCTCTCCAGGTAATTCATTAATTTGATGATCAGCTAAGAGGCCCATCTTTC 1224
 710 ACTCTCTCCAGGTAATTCATTAATTTGATGATCAGCTAAGAGGCCCATCTTTC 651

1225 TCACCTCTTCCGATGATGATATTCAGCAATATTTATTTAGACCCCTTACTGTGGGCAAA 1284
 650 TCACCTCTTCCGATGATGATATTCAGCAATATTTATTTAGACCCCTTACTGTGGGCAAA 591

1285 TCATTTGACGAGTAATTTGGAATAATGATATTCCTTATTCAGTAATGTCTACTGA 1344
 590 TCATTTGACGAGTAATTTGGAATAATGATATTCCTTATTCAGTAATGTCTACTGA 531

1345 GCACAACTAGTAATCATTAACGATAGGCTCATTTGTTTGAAGGTGTTATTTCA 1404
 530 GCACAACTAGTAATCATTAACGATAGGCTCATTTGTTTGAAGGTGTTATTTCA 471

1405 TAACAATATTTTACACCAATTCGATCAATGATATTTAGAACACATATATCAAGGA 1464
 470 TAACAATATTTTACACCAATTCGATCAATGATATTTAGAACACATATATCAAGGA 411

1465 TAAGTAATTTGAGGTATTCGCAATTTAAAGTATCCGATATTGATCACTATTATTA 1524
 410 TAAGTAATTTGAGGTATTCGCAATTTAAAGTATCCGATATTGATCACTATTATTA 351

1525 AATAATGAAAAAATGATTAATCTGTAATAAAGTGTATTTGTGACAGTCACTAATAT 1584

Db 350 AATAATGAAAAAATGATTAATCTGTAATAAAGTGTATTTGTGACAGTCACTAATAT 291

QY 1585 ACTAGATTTATTAATAATTTGTTTACTCTGCTCACCACAAACAGATGATATAACCC 1644

Db 290 ACTAGATTTATTAATAATTTGTTTACTCTGCTCACCACAAACAGATGATATAACCC 231

QY 1645 CAAATTAAGTATTAATCTTGATTAAGTATTAAGAGAGCTGGGTCTATTAATTAATTA 1704

Db 230 CAAATTAAGTATTAATCTTGATTAAGTATTAAGAGAGCTGGGTCTATTAATTAATTA 171

QY 1705 TTTTGAGGACAGACAGAGAGCTGTATTCCTAATGATTTAGTATGTTCTGTAATGAGAA 1764

Db 170 TTTTGAGGACAGACAGAGAGCTGTATTCCTAATGATTTAGTATGTTCTGTAATGAGAA 111

QY 1765 ATGTTACCAATTAATTTTACTTTTACTGATTTACATGATCATTTTATGAGGACATGTTCT 1824

Db 110 ATGTTACCAATTAATTTTACTTTTACTGATTTACATGATCATTTTATGAGGACATGTTCT 51

QY 1825 GGTATTAAGCAATTAATCTTTTATGATCA 1857

Db 50 GTGATTAAGCAATTAATCTTTTATGATCA 18

RESULT 12
 BM921276 1038 bp mRNA linear EST 12-MAR-2002
 LOCUS BM921276
 DEFINITION AGNCOURT_6626202 NIH_MGC_115 Homo sapiens CDNA clone IMAGE:5752662
 VERSION BM921276
 KEYWORDS BM921276.1 GI:19371655
 SOURCE EST.
 ORGANISM Homo sapiens (human)
 Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 1038)
 NIH-MGC http://mgc.nci.nih.gov/
 National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: c9abds-remail.nih.gov
 Tissue Procurement: Life Technologies, Inc.
 CDNA Library Preparation: Life Technologies, Inc.
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
 DNA Sequencing by: Agencourt Bioscience Corporation
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LNL at:
 http://image.lnl.gov
 Plate: LLM42787 row: d column: 07
 High quality sequence stop: 728.

FEATURES

Location/Qualifiers
 1..1038
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:5752662"
 /lab_host="DH10B"
 /clone_11b="NIH_MGC_115"
 /note="Organ: pooled brain, lung, testis; Vector:
 pCMV-SPORT6; Site 1: NotI; Site 2: EcoRV (destroyed); RNA
 source anonymous pool of 6 male brains, age range 23-27; 1
 male lung, age 27; and 1 male testis, age 69. Library is
 oligo-dT primed and directionally cloned (EcoRV site is
 destroyed upon cloning). Average insert size 1.8 kb,
 insert size range 1-3 kb. Library is normalized and
 enriched for full-length clones and was constructed by C.
 Gruber (Invitrogen). Research Genetics tracking code
 021. Note: this is a NIH_MGC Library."

ORIGIN

Query Match 34.2%; Score 640; DB 12; Length 1038;
 Best Local Similarity 99.9%; Pred. No. 3.6e-140;

Matches	760;	Conservative	0;	Mismatches	0;	Indels	1;	Gaps	1;
QY	36	TGAGCCGCGACGCGTTTGTGAGCCCGTTAGTGCGCCGCGGACGACGCGCCGCCAT	95						
DB	8	TGAGCCGCGACGCGTTTGTGAGCCCGTTAGTGCGCCGCGGACGACGCGCCGCCAT	67						
QY	96	GTCGCCGCTACTGCGTCCCCCAACAGCTCTCTTTGTGACGAAAGCTGGCGGACGAC	155						
DB	68	GTCGCCGCTACTGCGTCCCCCAACAGCTCTCTTTGTGACGAAAGCTGGCGGACGAC	127						
QY	156	CAGGCTGGAAGACTGGCGGCGTGAATTTGCTGATAGTCTTATGATGATGATGAT	215						
DB	128	CAGGCTGGAAGACTGGCGGCGTGAATTTGCTGATAGTCTTATGATGATGATGAT	187						
QY	216	TCCACTGATTTTCACTGCGCGTCCAGAGAGATTTGCTTATGATGATGATGAT	275						
DB	188	TCCACTGATTTTCACTGCGCGTCCAGAGAGATTTGCTTATGATGATGATGAT	247						
QY	276	TCTGATGCTGAAGACGCTTTACATATTTGGACAGAAAGTGGTGGACGCGACAT	335						
DB	248	TCTGATGCTGAAGACGCTTTACATATTTGGACAGAAAGTGGTGGACGCGACAT	307						
QY	336	TGAATACAGTTTGGCCAGGCGGATCGAAAGACCAATGATGAAAGCCAAAGAGG	395						
DB	308	TGAATACAGTTTGGCCAGGCGGATCGAAAGACCAATGATGAAAGCCAAAGAGG	367						
QY	396	GAGGAATGTGACAGTTCTTCACTGCTATGATGATGATGATGATGATGATGAT	455						
DB	368	GAGGAATGTGACAGTTCTTCACTGCTATGATGATGATGATGATGATGATGAT	427						
QY	456	CCGAAGTTATGAAGAGAGAGATCAAGAGTGGCTTTGATTAACAATATGAAGATC	515						
DB	428	CCGAAGTTATGAAGAGAGAGATCAAGAGTGGCTTTGATTAACAATATGAAGATC	487						
QY	516	GATATGCTCTAGAAACAGTACAGCACTGGAAGACCAAGCGGCTAG-AGAAGCATTCGA	574						
DB	488	GATATGCTCTAGAAACAGTACAGCACTGGAAGACCAAGCGGCTAG-AGAAGCATTCGA	547						
QY	575	CATGATACCAAACTGAGCTGGAATVCCAAGTACATTTCTGTTACTTACATTTCAAG	634						
DB	548	CATGATACCAAACTGAGCTGGAATVCCAAGTACATTTCTGTTACTTACATTTCAAG	607						
QY	635	AAAGATCTGAAGCGGAAAGAAACCAAGAGGAGGAGTTCAAGCGACCAAGGCTGCT	694						
DB	608	AAAGATCTGAAGCGGAAAGAAACCAAGAGGAGGAGTTCAAGCGACCAAGGCTGCT	667						
QY	695	GGAAGGTGCGAGTATGATATCTGACGAATTTTGAAGTCTGCTGCTGAAAGATPAAA	754						
DB	668	GGAAGGTGCGAGTATGATATCTGACGAATTTTGAAGTCTGCTGCTGAAAGATPAAA	727						
QY	755	GAATGTTATCGAAACTACATGGAATPATTGAAGTCCCTTC	795						
DB	728	GAATGTTATCGAAACTACATGGAATPATTGAAGTCCCTTC	768						
RESULT 13									
LOCUS	BX408797	895 bp	mRNA	linear	EST 13-MAY-2003				
DEFINITION	BX408797 Homo sapiens FETAL BRAIN Homo sapiens cDNA clone								
ACCESSION	CS00F019YN02	5-PRIME, mRNA sequence.							
VERSION	BX408797								
KEYWORDS	BX408797.1	GI:30652838							
SOURCE	EST.								
ORGANISM	Homo sapiens (human)								
REFERENCE	Bkayvota; Melazzo; Chordata; Craniata; Vertebrata; Euteleostomi;								
AUTHORS	1 (bases 1 to 895)								
TITLE	Li, W.B., Gruber, C., Jessee, J. and Polayes, D.								
JOURNAL	Full-length cDNA libraries and normalization								
COMMENT	Unpublished (2001)								
	Contact: Genoscope								
	Genoscope - Centre National de Sequencage								

QY	798	GTTCGAAAGTAAAGCAATTTTGAAGCAATTAAGAAATTTCAACTTGTGGAAGAC	857						
DB	22	GTTCGAAAGTAAAGCAATTTTGAAGCAATTAAGAAATTTCAACTTGTGGAAGAC	81						
QY	858	TAAATCCCTAATATGAATAGTTTATGATTCATGAGTAAAGGCTCAATTAATTA	917						
DB	82	TAAATCCCTAATATGAATAGTTTATGATTCATGAGTAAAGGCTCAATTAATTA	141						
QY	918	TGGAAGCTAGAGTCTGAATATCAAGAGACAGCCATGCTCTTACAGTGGCTCTG	977						
DB	142	TGGAAGCTAGAGTCTGAATATCAAGAGACAGCCATGCTCTTACAGTGGCTCTG	201						
QY	978	TTGGTCTGCTCAAACTGAATTTGGGAGAAAGGTATGCTCAATTAAGTCCATT	1037						
DB	202	TTGGTCTGCTCAAACTGAATTTGGGAGAAAGGTATGCTCAATTAAGTCCATT	261						
QY	1038	TTTGGCATTAATGGCAAACTGCTTGTATTTTGGTGGCAAGTCTTCTGCTTAAT	1097						
DB	262	TTTGGCATTAATGGCAAACTGCTTGTATTTTGGTGGCAAGTCTTCTGCTTAAT	321						
QY	1098	CATTGCTTGTGGGATCTGTTTATTTACTTGTACCAAGAGGATTAATCTGT	1157						
DB	322	CATTGCTTGTGGGATCTGTTTATTTACTTGTACCAAGAGGATTAATCTGT	381						
QY	1158	CTTAATCTCTCTTCCAGAGTAAATTCGAATTAATTTGATCCAGCTAAGAGGCCA	1217						
DB	382	CTTAATCTCTCTTCCAGAGTAAATTCGAATTAATTTGATCCAGCTAAGAGGCCA	441						
QY	1218	TCTCTCTCAAGCTCTTCCAGAGTAAATTCGAATTAATTTGATCCAGCTAAGAGGCCA	1277						
DB	442	TCTCTCTCAAGCTCTTCCAGAGTAAATTCGAATTAATTTGATCCAGCTAAGAGGCCA	501						
QY	1278	GAGCAATCATTTGATGATTAATTAAGAAATTAATTAATTTGATCCAGCTAAGAGGCCA	1337						
DB	502	GAGCAATCATTTGATGATTAATTAAGAAATTAATTAATTTGATCCAGCTAAGAGGCCA	561						
QY	1338	CTAAGAGCAATCTAAGTAAATTAATTAAGAAATTAATTAATTTGATCCAGCTAAGAGGCCA	1397						
DB	562	CTAAGAGCAATCTAAGTAAATTAATTAAGAAATTAATTAATTTGATCCAGCTAAGAGGCCA	621						
QY	1398	TTAATTCATCAATATTTTACACCATTCGATCAATG	1434						
DB	622	TTAATTCATCAATATTTTACACCATTCGATCAATG	658						

ORIGIN

Query Match 34.0%; Score 637; DB 13; Length 895;
 Best Local Similarity 100.0%; Pred. No. 2e-139;
 Matches 637; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

FEATURES

source

BP 191 91006 EVRY cedex - France
 Email: segre@genoscope.cns.fr, Web: www.genoscope.cns.fr
 Library was constructed by Life Technologies, a division of
 Invitrogen. This sequence belongs to sequence cluster 798.r For
 more information about this cluster, see
http://www.genoscope.cns.fr/cgi-bin/cluster.cgi?seq=CS0BAA010ZC02_CS00876_1&cluster=798.r
 Contact: Feng Liang Email: fliang@lifetech.com URL:
<http://fulllength.invitrogen.com/invitrogen/1600>
 Faraday Avenue Genoscope sequence ID: CS0BAA010ZC02_CS00876_1.

Location/Qualifiers

1..895
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="CS00F019YN02"
 /tissue_type="FETAL BRAIN"
 /dev_stage="fetal"
 /clone_id="Homo sapiens FETAL BRAIN"
 /note="Organ: brain; Vector: pCMVSPORT 6; 1st strand cDNA
 was primed with a NotI-oligo(dT) primer. Five prime end
 enriched, double-strand cDNA was digested with Not I and
 cloned into the Not I and EcoRV sites of the pCMVSPORT 6
 vector. Library was not normalized."

	RESULT 14
	BUG621746/c
LOCUS	BUG621746
DEFINITION	UI-H-FLI-bga-p-10-0-UI.s1 NCI CGAP FLI Homo sapiens CDNA clone
ACCESSION	U1-H-FLI-Bga-p-10-0-UI 3', mRNA sequence.
VERSION	BU621746
KEYWORDS	BUG621746.1 GI:23287961
SOURCE	EST.
ORGANISM	Homo sapiens (human)
REFERENCE	Homo sapiens Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Hominidae; Homo. 1 (bases 1 to 670) NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap. National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
AUTHORS	Unpublished (1997)
JOURNAL COMMENT	Contact: Robert Strausberg, Ph.D. Email: cgsabbs-re@mail.nih.gov Tissue Procurement: James Martin CDNA Library Preparation: Dr. M. Bento Soares, University of Iowa CDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa DNA Sequencing by: Dr. M. Bento Soares, University of Iowa Clone Distribution: Clone distribution information can be obtained from Dr. M. Bento Soares, bento-soares@uiowa.edu The following repetitive elements were found in this CDNA sequence: 71..120, >LINE2 Seq primer: M13 FORWARD POLYA=yes.

FEATURES	Location/Qualifiers
source	1. .670

```

"organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="UT-H-Flu-bga-p-10-0-UT"
/tissue_type="Cell lines"
/dev_stage="Adult"
/lab_host="DH10B (Life Technologies)"
/clone_lib="NCI CGAP_Flu"
/note="Organ: Chondrosarcoma; Vector: pT773-Pac
Pharmacia) with a modified polylinker; Site 1: Bscr I,
Site 2: Not I; NCI CGAP Flu is a normalized cDNA library
derived from a pool of mRNA obtained from 4 cell lines
from grade III chondrosarcoma tissues. The library was
constructed according to Bonaldo, Lemon and Soares,
Genome Research, 6:791-806, 1996. First strand cDNA
synthesis was primed with an oligo-dT primer containing a
Not I site. Double stranded cDNA was ligated to an EcoR I
adaptor, digested with Not I, and cloned directionally
into pT773-Pac vector. The oligonucleotide used to prime
the synthesis of first-strand cDNA contains a library tag
sequence that is located between the Not I site and the
(dT)18 tail. The sequence tag for this library is
GAGGTCGGTG. The cell lines were provided by Dr. James
Martin from the University of Iowa.
TAG_TISSUE=human Chondrosarcoma Grade 3 cell line mix
TAG_LIB=UT-H-Flu
TAG_SEQ=GAGGTCGGTG"

```

Query Match	32.2%	Score 603	DB 13	Length 670
Best Local Similarity	99.8%	Pred. No. 2.1e-131		
Matches 653	Conservative	0	Mismatches 1	Indels 0
				Gaps 0

Qy	69	AAGGTCCTCAGATGAATATCTACGTACGAAATTTTGTGACCTCGGCTCAAAAAGATTAAGA	756
Db	670	AAGGTCCTCAGATGAATATCTACGTACGAAATTTTGTGACCTCGGCTCAAAAAGATTAAGA	611
Qy	757	AATGTTATCGAAACTACATGAAATATATGAAAGTCCCTTGAAGTTGAAAGTAAACAATTTT	816
Db	610	ATGTTATCGAAACTACATGAAATATATGAAAGTCCCTTGAAGTTGAAAGTAAACAATTTT	551

QY	817	AGCAAAATATAAGAAATTCACATTGTGACTTGCGGAAACCTAATCCCTAATATATGATA	876
Db	550	AGGCAAAATATAAGAAATTCACATTGTGACTTGCGGAAACCTAATCCCTAATATGATA	491
QY	877	GGTTTATATTGATTCATGGGTAAACAGGTCCATTAATTAATTGTGAAACTAGATGTCTG	936
Db	490	GGTTTATATTGATTCATGGGTAAACAGGTCCATTAATTAATTGTGAAACTAGATGTCTG	431
QY	937	AATATCAAAGAAAGACAGCATAGCTCTTAACAGGCGCTGTGGTCTGTCTCAACACGA	996
Db	430	AATATCAAAGAAAGACAGCATAGCTCTTAACAGGCGCTGTGGTCTGTCTCAACACGA	371
QY	997	ATTGGGTGGGAAAAGTATGGTCCAAATAAAGTTCATTTTGGCTTAATTGGCAAT	1056
Db	370	ATTGGGTGGGAAAAGTATGGTCCAAATAAAGTTCATTTTGGCTTAATTGGCAAT	311
QY	1057	CTTGCGTTTGGTTATTTTGGTGGCAAGTCTTTCGTAACTAATTTGGTTGGTCATC	1116
Db	310	CTTGCGTTTGGTTATTTTGGTGGCAAGTCTTTCGTAACTAATTTGGTTGGTCATC	251
QY	1117	TGTGTTATTTACTGTGTACACCAATCAGTTTACATCTGTCTTAACACTCCTCCGAG	1176
Db	250	TGTGTTATTTACTGTGTACACCAATCAGTTTACATCTGTCTTAACACTCCTCCGAG	191
QY	1177	GTAATTCGCAATATATTTGACATCCGAGCTTAAGAGGCGCATCTCTTCTACCTCTTCC	1236
Db	190	GTAATTCGCAATATATTTGACATCCGAGCTTAAGAGGCGCATCTCTTCTACCTCTTCC	131
QY	1237	TAGTCAGTATATTCAGCAAAATATTTATTTGAGCCCTTACTGTGGCAATCATTTGACTGG	1296
Db	130	TAGTCAGTATATTCAGCAAAATATTTATTTGAGCCCTTACTGTGGCAATCATTTGACTGG	71
QY	1297	ATAATTGAGAAAAATATATTAATTTCCCTTATTCAGTAATGTCTTACTGACGACAA	1350
Db	70	ATAATTGAGAAAAATATATTAATTTCCCTTATTCAGTAATGTCTTACTGACGACAA	17

RESULT	15
Locus	BM728917
Definition	760 bp mRNA linear EST 01-MAR-2002 UT-E-BOL-aiv-h-08-0-UT.r1 UT-E-BOL Homo sapiens cDNA clone UT-B-BOL-aiv-h-08-0-UT.5', mRNA sequence.
Accession	BM728917
Version	BM728917.1 GI:19050250
Keywords	EST,
Source	Homo sapiens (human)
Organism	Homo sapiens
Reference	Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo. 1 (bases 1 to 760) Bonaldo,M.F., Lennon,G. and Soares M.B. Normalization and subtraction: two approaches to facilitate gene discovery
Journal	Genome Res. 6 (9), 791-806 (1996)
Medline	97044477
PubMed	8889548
Comment	Contact: Soares, MB

FEATURES	Location/Qualifiers
source	1. .760

1. .760

```
/organism="Homo sapiens"  
/mol_type="mRNA"  
/db_xref="taxon:9606"  
/clone="UI-E-EO1-a1v-h-08-0-UI"  
/tissue_type="fetal eye"  
/dev_stage="fetal"  
/lab_host="DH10B (Life Technologies) (T1 phage resistant)"  
/clone_11b="UI-E-EO1"  
/note="Organ: eye; Vector: pT73-Pac (Pharmacia) with a  
modified polylinker; Site_1: EcoR I; Site_2: Not I;  
UI-E-EO1 is a normalized cDNA library containing the  
following tissue(s): fetal eye. The library was  
constructed according to Bonaldo, Lennon and Soares,  
Genome Research, 6:791-806, 1996. First strand cDNA  
synthesis was primed with an oligo-dT primer containing a  
Not I site. Double stranded cDNA was ligated to an EcoR I  
adaptor, digested with Not I, and cloned directionally  
into pT73-Pac vector. The oligonucleotide used to prime  
the synthesis of first-strand cDNA contains a library tag  
sequence that is located between the Not I site and the  
(dT)18 tail. The sequence tag for this library is  
CCGATATACC. This library was created for the program, Gene  
Discovery in the Visual System, supported by National Eye  
Institute (NEI)."
```

ORIGIN

```
Query Match      31.8%; Score 596; DB 12; Length 760;  
Best Local Similarity 100.0%; Pred. No. 8.1e-130; Indels 0; Gaps 0;  
Matches 596; Conservative 0; Mismatches 0;
```

```
QY 582 AGACCAACTGCAAGTGAATACCCAGTACAGTTCTGTTACTACACTTCAAGAAAGATC 641  
|  
|  
|  
Db 44 AGACCAACTGCAAGTGAATACCCAGTACAGTTCTGTTACTACACTTCAAGAAAGATC 103  
|  
|  
|  
QY 642 TGAAGCGGAAAAAGAACCAAGAGGGGAGTTCAAGCAACCAAGGGTGGTGAAGGT 701  
|  
|  
|  
Db 104 TGAAGCGGAAAAAGAACCAAGAGGGGAGTTCAAGCAACCAAGGGTGGTGAAGGT 163  
|  
|  
|  
QY 702 GCTGCGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 761  
|  
|  
|  
Db 164 GCTGCGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 223  
|  
|  
|  
QY 762 ATCGAAACTACATGGAATTAATGAATGATGATGATGATGATGATGATGATGATGAT 821  
|  
|  
|  
Db 224 ATCGAAACTACATGGAATTAATGAATGATGATGATGATGATGATGATGATGATGAT 283  
|  
|  
|  
QY 822 AATTAAGGAAATTAATGATGATGATGATGATGATGATGATGATGATGATGATGAT 881  
|  
|  
|  
Db 284 AATTAAGGAAATTAATGATGATGATGATGATGATGATGATGATGATGATGATGAT 343  
|  
|  
|  
QY 882 AATTAAGGAAATTAATGATGATGATGATGATGATGATGATGATGATGATGATGAT 941  
|  
|  
|  
Db 344 AATTAAGGAAATTAATGATGATGATGATGATGATGATGATGATGATGATGATGAT 403  
|  
|  
|  
QY 942 CAAGGAAGACAGCATAGTCTTAACTGATGATGATGATGATGATGATGATGATGATGAT 1001  
|  
|  
|  
Db 404 CAAGGAAGACAGCATAGTCTTAACTGATGATGATGATGATGATGATGATGATGATGAT 463  
|  
|  
|  
QY 1002 GTGGGAAAAAGGTATGTCATATTAAGTTCATTTTCCATTAATGCGCAAAATCTTGC 1061  
|  
|  
|  
Db 464 GTGGGAAAAAGGTATGTCATATTAAGTTCATTTTCCATTAATGCGCAAAATCTTGC 523  
|  
|  
|  
QY 1062 CTTTGTATTTTGTGGCGAGTGTTCCTTAATCATTTGCTTGTGTCATGTCATGTCAT 1121  
|  
|  
|  
Db 524 CTTTGTATTTTGTGGCGAGTGTTCCTTAATCATTTGCTTGTGTCATGTCATGTCAT 583  
|  
|  
|  
QY 1122 TTATTTACTGTATACCAATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1177  
|  
|  
|  
Db 584 TTATTTACTGTATACCAATGATGATGATGATGATGATGATGATGATGATGATGATGAT 639  
|  
|  
|
```

```
Search completed: July 29, 2004, 09:21:48  
Job time : 5004 secs
```